INFORMATION-SELECTIVITY OF ALZHEIMER'S DISEASE PROGRESSION

by

MARK STEPHEN ROWAN

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Computer Science
College of Engineering and Physical Sciences
The University of Birmingham
April 2013

For Frances,

who inspired

this work.

ACKNOWLEDGEMENTS

Numerous people helped me throughout the journey which constituted the writing of this thesis – and this is the part where I get to thank them.

Firstly, my supervisor, John Bullinaria, whose gentle guidance (and occasional contributory head-scratching) gave me the freedom to learn and explore what, to both of us, was a new field. Even when things seemed to be going down a dead-end, John could find something for us to laugh about, then calmly suggest a new direction. (As my MSc and BSc supervisor as well, it's a wonder he didn't get fed up of me – but working together for eight years has been productive and, above all, fun – so thank you!) Also Jon Rowe and Xin Yao, for their comments and contributions during thesis group meetings, and John Jefferys from the School of Neuroscience for offering additional advice.

Sam Neymotin and Bill Lytton from SUNY Downstate, Brooklyn, with whom I collaborated on the spiking neural network experiments, for their help and patience when debugging the model and trying to understand the code; and Marshall Crumiller from Mount Sinai School of Medicine, New York, for his help with getting the Fourier information calculation code running. Thanks also to the anonymous reviewers of my papers.

A PhD may ultimately be a solo journey, but it never has to be lonely. So thank you to my friends and colleagues in Computer Science – in particular, the past and present occupants of Room 117, and Vivek Nallur, for the various discussions and activities we enjoyed, which took my mind off the PhD for long enough to remember to have fun.

Finally, biggest thanks of all go to Hannah who, during the course of my PhD, on top of a mountain in the rain in Zürich, made me incredibly happy by agreeing to be my fiancée. You have helped me in more ways than you can know :)

Abstract

Alzheimer's disease (AD) is a growing global healthcare problem, as life expectancy increases and populations age. Current treatments focus on reducing symptoms, rather than treating the underlying causes of the disease, and as such are disappointing in their efficacy. One reason for this is the current poor understanding of the mechanisms of disease pathology.

An existing hypothesis of AD progression predicts that homeostatic synaptic scaling mechanisms, which normally act to balance potentiation during learning, may also direct the progression of the disease throughout the brain as cells scale up their sensitivity to compensate for lost activation. This thesis makes the additional prediction that such a mechanism would be likely to target those cells with the lowest contribution of information to the network in early stages of the disease, resulting in the delayed onset of cognitive symptoms and making timely intervention and treatment of the disease more difficult.

A computational modelling approach is used to investigate these hypotheses, firstly using an existing abstract Hopfield-type neural network. The model was extended to incorporate homeostatic synaptic scaling, and information theoretic measures were used to characterise the information contribution of individual cells within the network. The model was then lesioned according to the scaling-driven progression hypothesis of AD, showing that the pathology is capable of targeting neurons with the lowest information contribution to the network at early stages of the disease, and therefore resulting in a delayed onset of cognitive symptoms. Additional experiments revealed a positive-feedback loop by which noisy compensatory synaptic scaling mechanisms caused the accelerated degradation of recent memories, which were themselves preferentially used as drivers of the compensatory mechanism.

The hypothesis was then tested in a biologically-realistic spiking model of neocortex, which was also extended to include the effects of synaptic scaling and to operate on very

long simulated timescales. A study was undertaken to demonstrate the effects of the interaction between synaptic scaling and potentiation during learning. A recent method for obtaining mutual information between cells in large networks of spiking neurons, based on Fourier analysis of spike times, was applied to the model, and a regime of stimulation was developed to elicit reliable information measures. Cell death, modelled as an abstract excitotoxicity mechanism based on scaling factor values, confirmed the earlier results and showed that low-information neurons (and neurons from cortical layers with the lowest information contribution) were the first to die in scaling-driven AD pathology. Another (not validated) biophysical mechanism for excitotoxicity revealed that, although the effect of information-selectivity was maintained, the order of deletion was reversed, with high-information cells dying first. However, other features of this mechanism produced biologically implausible results, suggesting that the true identity of biological toxicity mechanisms may be different.

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Motivation

Alzheimer's disease (AD) is a specific form of dementia, characterised biologically by neurofibrillary tau protein tangles and beta-amyloid (Aβ) protein plaques (Tiraboschi et al., 2004), and symptomatically by a progressive decline in memory capabilities, leading to a loss of personality and motor function, and eventually to death. Recent health reports have given dire warnings of future Alzheimer's disease epidemics in coming years due to the aging of world populations (Mount and Downton, 2006). Predictions of the tripling of cases of AD within the next few decades (Minati et al., 2009), if they come to pass, will result in a huge strain on healthcare budgets, estimated at £23bn per year for the UK economy alone (Boche and Nicoll, 2010). Yet despite the massive implications of this rising epidemic, little is known today about the causes of, or potential cures for, Alzheimer's disease. The main tools currently available to doctors are drugs that simply prolong the disease by mitigating cognitive symptoms, rather than stopping the disease altogether.

Computational modelling of neurological disorders such as AD is an established tool for research into the understanding of neurological disorders, and as a way of making testable predictions which are abstracted from the uncertainties of experimentation on real tissue in a laboratory (Aakerlund and Hemmingsen, 1998). Connectionist models

typically use recurrent networks (Willshaw et al., 1969; Hopfield, 1982) to model the way in which associative memory in cortical microcolumns behaves, but these simplified artificial neural networks are not intended to be used to simulate the exact processes occurring in the brain; instead, they aim to abstract out the workings of low-level network behaviours and create a computational analogy which can be used to test specific theories about how the brain works (Duch, 2007).

Existing connectionist models of AD (Ruppin and Reggia, 1995a; Hasselmo, 1994) can now be made capable of better approximation of current medical knowledge. Today, more is known about the underlying pathology such as tau neurofibrillary tangles (Ballatore et al., 2007; Spires-Jones et al., 2009; Feinstein and Wilson, 2005) and amyloid plaques (Hardy and Selkoe, 2002; Hardy, 2009), with hypotheses such as the synaptic scaling-driven progression of AD (Small, 2008) leading to new directions for research. Additionally, more biologically-realistic network architectures such as reservoir networks (Yamazaki and Tanaka, 2007; Lukosevicius and Jaeger, 2009), networks of spiking neurons (Gerstner and Kistler, 2002), and learning rules such as LEABRA (O'Reilly, 2001) exist, which could be incorporated into connectionist models to improve the accuracy and range of their predictions.

One current research area of crucial importance results from the fact that cognitive symptoms of AD do not become apparent until relatively late in the progression of the disease. This makes timely intervention more difficult, and subsequent treatment less effective. Therefore, much current research focuses on earlier detection of the presence of AD using indirect methods such as EEG analysis of alpha rhythms (Bhattacharya et al., 2011), rather than waiting for the onset of cognitive decline. By better understanding the mechanisms of disease progression, it may be possible to help direct the search for such bio-markers of disease presence. Additionally, it may even be possible to propose novel treatments which operate directly on the progression mechanisms of AD rather than simply trying to mitigate the progression of cognitive symptoms.

This thesis therefore aims to improve on historical connectionist models of AD by

incorporating recent medical knowledge and theories of progression, to better understand the mechanisms of AD at the network level. This should lead to testable predictions, and the possibility of identifying novel treatments to slow or halt the progression of the disease.

1.2 Research questions

Throughout the progress of this work, a few fundamental questions were key to directing the research. Some of these, once investigated, led in turn to new questions and hypotheses. The following are the three major questions which were posed at the very beginning of the research period:

- What improvements can be made to existing connectionist models of Alzheimer's disease, in terms of biological and computational advancements?
- Why do cognitive symptoms of Alzheimer's disease appear so late after onset of the disease?
- How might we detect Alzheimer's disease before the appearance of cognitive symptoms, or how might we attempt to prevent cognitive decline before the appearance of symptoms?

These over-riding questions led to further, more specific questions which were investigated during the research period:

- What effects do properties such as connectivity and capacity have on the progression of Alzheimer's disease?
- How does compensatory synaptic scaling affect memory maintenance?
- How does a neuron know what firing rate it should maintain during scaling?
- What effects does synaptic scaling have on learning?

- How can we measure the significance of a neuron in a noisy spiking neural network?
- Are low-significance neurons more susceptible to death at the start of Alzheimer's disease, potentially explaining the delayed onset of symptoms?
- Can low-intensity neurostimulation act to reduce instability resulting from compensatory synaptic scaling, and slow the progression of Alzheimer's disease?

1.3 Contributions to knowledge

The following summarises the major contributions to knowledge contained within this thesis. The individual items are explained in more detail in the relevant chapters, and again in the evaluation at the end of the thesis (chapter 8).

Compensation-driven damage The identification and study of a positive feedback loop of noisy synaptic compensation, caused by – and further causing – the accelerated degradation of recent memories in Alzheimer's disease, which are themselves preferentially used as drivers of a compensatory mechanism (chapter 3).

Information-selectivity hypothesis The formulation of a theory of information-selectivity of Alzheimer's disease pathology, in which neurons with the least significance to the network are selectively targeted first. An abstract computational model is developed, showing that information-selectivity gives rise to a delayed onset of cognitive symptoms and hides the presence of the disease, making timely intervention more difficult (chapter 4).

Periodic-update activity sensor The modification of van Rossum et al. (2000)'s slow-varying activity sensor, which allows a neuron to vary its sensitivity in order to attain its target firing rate, to work with periodic updates instead of fixed timesteps. This makes it possible to use the activity sensor with modern fast just-in-time neural simulations, rather than being limited to simulations with a fixed timestep.

Synaptic scaling and learning interactions A clear demonstration of the effects of synaptic scaling on balancing hyperactivity during learning is provided, using a biologically-realistic cortical columnular model, and operating for the first time on realistic timescales of "hours to days". Additionally, the stability of the synaptic scaling mechanism is demonstrated to be disrupted during Alzheimer's-like damage, giving rise to transient epileptiform bursting, matching similar medical observations in AD patients (chapter 6).

Information contribution in spiking networks The application of a recent method for obtaining the mutual information between all cells in a group (Crumiller et al., 2011) to a biologically-realistic spiking model of neocortex. This leads to the development of a regime of network stimulation which successfully probes the synaptic weights, and reveals levels of information contribution using the Crumiller method (chapter 7).

Information variations in cortical column An explicit demonstration and quantification of the differences in information contribution across different layers of the neocortical column, matching predictions by Neymotin et al. (2011c) (chapter 7).

Compensatory down-scaling during network damage The identification of a mechanism by which a significant proportion of cells may also have to scale *down* rather than *up* during compensatory scaling processes, thereby modifying the hypothesis of Small (2008) (chapter 7).

Relationship between cell death and information contribution Confirmation, in a biologically-realistic neocortical column model, of the existence of a relationship between information contribution and cell death during synaptic scaling-driven pathology in simulated Alzheimer's disease (chapter 7).

Information-selectivity in Alzheimer's disease The presentation of results showing that an abstracted toxicity mechanism which causes cell death in proportion to

scaling factor values, as predicted by Small (2008, 2009), is capable of selectively targeting cells with low information-contribution to the network in a biologically-realistic computational model, thus providing a mechanism for the delayed onset of cognitive symptoms in AD which makes timely intervention and treatment more difficult (chapter 7).

1.4 Methodology

The research followed an open-ended exploratory and experimental methodology to enable the development and subsequent empirical testing of hypotheses. The research began using the three major questions listed earlier as a guide for this process.

An investigation into current medical theories and mechanistic processes of Alzheimer's disease was undertaken, and previous connectionist models of AD were analysed to identify areas in which they could be brought into line with current disease progression theories. This led to the development of new hypotheses which were outlined and tested using the connectionist models as examples.

During the course of the research, it became clear that it would be necessary to test these hypotheses in a more biologically-realistic manner in order to verify their predictions. A full examination of the literature was undertaken to identify the required features of a biologically-realistic model which could support the testing of the hypotheses. A comparison of a number of published, biologically-validated spiking neural models was made against these criteria, in order to select a suitable candidate model. The model was analysed and the hypotheses were tested experimentally.

1.5 Thesis outline

This chapter (chapter 1) introduces the thesis and presents the over-riding motivation for the research, along with the questions that were used to guide the research.

Chapter 2 gives an in-depth background to the key current medical theories of AD, including the synaptic scaling-driven progression hypothesis (Small, 2008) on which later parts of this thesis are based, and a review of the field of computational modelling of neurological disorders.

Chapter 3 introduces a Hopfield-type model of memory decline in Alzheimer's disease with compensatory synaptic scaling (Ruppin and Reggia, 1995a), and experimentally characterises it to describe ways in which the model can be brought into line with current medical knowledege. This work leads to suggested therapeutic strategies which may reduce the progression of damage.

The synaptic scaling-driven progression hypothesis of AD is examined in closer detail in chapter 4 in the context of the Ruppin and Reggia (1995a) model, and a hypothesis of the selectivity of AD for low-information neurons at early stages of the disease is developed. This hypothesis could provide an explanation for the delayed appearance of cognitive symptoms after the onset of disease pathology.

Next, chapter 5 lays out the necessary considerations for testing the hypothesis of information-selectivity within biologically-realistic spiking neural models, which is vital for making biologically-testable predictions of disease mechanisms and for proposing new treatments. A survey of existing biologically-validated models is undertaken, and a strategy for obtaining estimations of neural information contribution is outlined. A discussion of the open problem of defining neural activity targets is presented.

Chapter 6 examines synaptic scaling in a selected biologically-realistic spiking neural model of a neocortical column, showing its effects on balancing potentiation during learning to prevent hyper-potentiation, and in compensating for lost activation during network damage. Experimental results are presented to show that synaptic scaling does not interfere with learning.

Chapter 7 then introduces the synaptic scaling-driven progression mechanism of AD pathology into the spiking model. A regime of network stimulation which elicits valid information calculations is experimentally identified, and the information-selectivity hy-

pothesis described in chapter 4 is tested to observe its validity in a biologically-realistic neocortical model.

Finally, chapter 8 provides a critical evaluation and a summary of the work undertaken for this thesis, as well as proposing future work directly resulting from this research, including a hypothesis for a possible novel treatment of AD.

1.6 Resulting publications

The following conference and journal articles have appeared as a direct result of research undertaken during the preparation of this thesis:

- M. Rowan. Effects of Compensation, Connectivity and Tau in a Computational Model of Alzheimer's Disease. In *The 2011 International Joint Conference on Neural Networks (IJCNN)*, pages 543–550. IEEE, 2011

- M. Rowan. Information-selectivity of beta-amyloid pathology in an associative memory model. *Frontiers in Computational Neuroscience*, 6(2), January 2012

- M. Rowan and S. Neymotin. Synaptic scaling balances learning in a spiking model of neocortex. In M. Tomassini et al., editor, *11th International Conference on Adaptive and Natural Computing Algorithms (ICANNGA)*, volume 7824 of *Lecture Notes in Computer Science*, pages 20–29, Lausanne, Switzerland, 2013. Springer-Verlag

CHAPTER 2

BACKGROUND

Chapter synopsis

There are many competing and complementary hypotheses of the mechanisms of Alzheimer's disease. Some of these attempt only to explain the observed cognitive deficit symptoms; others provide a more complete biochemical mechanism. The focus of this work is on connectionist modelling rather than biochemical simulation, so it does not aim to deal with the complexities of modelling cellular-level alterations to the various signalling pathways.

This chapter introduces the medical background to Alzheimer's disease, including the various hypotheses regarding the pathologies which occur in AD. A summary of these is given alongside suggested methods of representative computational lesioning.

A general overview is then given of the field of computational modelling of the brain using artificial neural networks. We introduce two classes of simple associative neural networks – the Hopfield model and the Willshaw model – as well as some biologically-inspired learning rules which can be used to train these networks. The notion of sparse connectivity is introduced, as well as the concept of networks of spiking neurons, which shall be useful in later chapters. Some of the shortfalls of basic associative networks are addressed, with a potential solution for these problems outlined in an introduction to reservoir computing. Finally, some general advice is given on abstraction and over-simplification in computational modelling.

2.1 Alzheimer's disease

2.1.1 Glossary

The following is intended as a brief glossary of medical and biological terms which are relevant to the computational study of mechanisms of Alzheimer's disease. Knowledge of these terms will be of benefit in later chapters, particularly those dealing with biologically-realistic modelling. The scope of this glossary is neither broad nor deep, but serves to give the minimum amount of background knowledge needed to understand the concepts presented in this thesis.

Hippocampus The region of the brain central to episodic and associative memory.

Neocortex Much more recent region of the brain (in evolutionary terms) which deals with higher-level processing.

Cortical column Postulated structure of the cortex, consisting of neurons in a number of discrete layers exhibiting both feed-forward connectivity between layers, and recurrent connectivity throughout.

Pyramidal neuron Excitatory neuron found in the cortex and hippocampus.

Inhibitory interneuron Neurons which use GABAergic firing via Cl^- ions to inhibit the firing of postsynaptic neurons.

Oscillatory rhythms Emergent oscillations in biological neural networks, banded according to their frequencies, with delta rhythms at 1-4Hz, theta at 4-8 Hz, alpha at 8-12 Hz, beta at 12-30 Hz, and gamma at 30-70 Hz.

STDP Spike Time-Dependent Plasticity; a biological learning rule which strengthens synapses in a Hebbian manner when two neurons fire spikes in close correlation with each other, and depresses them otherwise.

β-amyloid Core component of amyloid plaques, found throughout the brains of patients with Alzheimer's disease. Whether β-amyloid plaques are a toxic cause of AD, or a result of some other process, is not currently known.

Tau Protein essential for a neuron's structural support and function, but in AD is hyperphosphorylated and sequestered into tangles. These neurofibrillary tangles (NFTs) lead to the collapse of normal tau structural support, as well as physically blocking neurotransmitter vesicles.

Necrosis Traumatic cell death, in which a cell ceases to function, its membrane breaks down, and parts of the dead cell are leaked into the intracellular space, causing an inflammatory response in the rest of the network and the spread of further damage.

Apoptosis Programmed cell death, e.g. in response to excessive calcium influx in order to safely destroy a cell and prevent necrosis, or as part of neural network development, e.g. pruning of unnecessary connections and neurons.

Neurotrophin A protein which signals to neurons to prevent apoptosis, and induces them to grow during neural development stages.

Brain-Derived Neurotrophic Factor (BDNF) A specific neurotrophin secreted in response to activity in the brain, which enhances inhibitory signalling as part of a global activity balancing mechanism. Reduced BDNF has the effect of increasing pyramidal cell excitation (Turrigiano, 2008).

Tumor necrosis factor-alpha (TNF-α) A protein released by supporting *glial cells* in the brain, as a response to lowered activity. TNF-α enhances excitatory signalling as part of a global activity balancing mechanism (Turrigiano, 2008).

Glutamate Excitatory neurotransmitter group causing Ca^{2+} or Na^{+} ion influx.

AMPA Receptor for a specific glutamate subtype, permitting excitatory Na^{+} ion uptake.

NMDA Receptor for a specific glutamate subtype, permitting excitatory Ca^{2+} uptake.

GABA An inhibitory neurotransmitter, which causes Cl^- influx into the cell, thus dampening the neuron's response to excitation. GABA receptors comes in two forms ($GABA_A$ and $GABA_B$) with differing speeds of activation, thus enhancing the dimensionality of inhibitory dynamics.

Acetylcholine (ACh) Neurotransmitter which "clamps" activation of previously-stored patterns during associative learning. Degeneration of ACh-releasing neurons is implicated in cognitive decline seen in Alzheimer's disease.

2.1.2 Cholinergic hypothesis

Loss of cholinergic neurons

The earliest hypothesis for the mechanism of Alzheimer's disease was proposed by researchers studying the effects of the neurotransmitter acetylcholine (ACh) and its role in learning and human memory (Drachman and Leavitt, 1974; Bartus et al., 1982). The theory states that cell death occurring in regions of the brain which produce ACh causes "a loss of cholinergic function in the central nervous system [which] contributes significantly to the cognitive decline associated with advanced age and AD" (Terry and Buccafusco, 2003). This hypothesis is supported by a range of evidence from human and animal observations in addition to the generally positive response observed in AD sufferers after supplying acetylcholinesterase inhibitors (AChEIs). These are drugs designed to block the action of acetylcholinesterase enzymes (AChE) which break down ACh, leading to a greater level of remaining ACh in the brain (Buccafusco and Terry, 2000), and therefore a slowing of cognitive decline. Conversely, localised application of cholinergic antagonists (cholinergic function-reducing substances) in animal experiments has shown a subsequent decline in cognitive function (Hasselmo, 2006).

Role of acetylcholine

Spitzer (1998) in Stein and Ludik (1998) explains that "during learning, the spreading of activation through the [neural] network must be prevented to some degree, otherwise synaptic activation and change would spread like an avalanche through the entire network [...] Acetylcholine selectively suppresses excitatory synaptic connections between neurons of the same cortical region. In contrast, signals from other cortical areas can pass through synapses unimpaired".

A computational analogy of this function of acetylcholine is to 'clamp' the activation of neurons in a neural network during learning, to prevent the interfering effect of spontaneously recalled previously-stored overlapping patterns (Hasselmo, 1993). Without this clamping effect of ACh, "when a new memory pattern is being stored in the network, the resulting network activity is not only guided by the new pattern but also by all the previous memory patterns which are engraved in the synaptic matrix. Thus, previously memorized patterns tend to bias the activation during new storage in 'their direction'. This inherent reinforcement may lead to exponential synaptic growth and to a pathological increase in the number of synapses" (Ruppin, 1995), and also potentially to neuronal death via over-excitation of the cell.

This process has been termed *runaway synaptic modification* (Hasselmo, 1994), in which reduced cholinergic suppression of the activation of previously-stored patterns during learning leads to an exponential increase in the number of synapses in the network. This causes excess activity in unrelated regions during recall, as well as causing the network to become over-sensitive to noise, and leads to an overall cognitive decline. However, the results seen in this model (exponential synaptic growth and a resulting pathological increase in the number of synapses) are inconsistent with the results of other contemporary work such as Ruppin and Reggia (1995a), and in general in Alzheimer's diseasae, in which reduced synaptic density is predicted.

A further model of ACh action in the CA3 region of the hippocampus shows that the role of acetylcholine in cellular-level regulation may be to switch the network between one

of two firing modes ('burst' and 'regular spiking'), related to a transition between fast learning and consolidation of memory Menschik and Finkel (1999). Reduced acetylcholine levels translated into a reduced time allowed for the attractor network to settle into the appropriate basin of attraction for a given input, leading to an increase in the retrieval failure rate of the network.

Problems with the cholinergic hypothesis

The cognitive effects of reduced cholinergic signalling in AD is well-supported, and for this reason the cholinergic hypothesis has been the major drive for drug development in AD until recently (Minati et al., 2009). AChEIs remain the most widely prescribed drugs for treatment of AD, yet the cholinergic hypothesis as it stands does not explain either the formation of β-amyloid plaques or of neurofibrillary tau protein tangles observed in the brains of patients with Alzheimer's disease, but only the presence of the observed cognitive deficit symptoms when ACh signalling mechanisms are disrupted (Terry and Buccafusco, 2003). "Because of the main mechanism of action, therapy with cholinesterase inhibitors is not expected to significantly alter the accumulation of underlying pathology; it can only temporarily mitigate symptoms" (Minati et al., 2009).

The level of efficacy of AChEIs in preventing symptomatic regression is debated, and better results have been observed in early trials of β-amyloid immunotherapy drugs (Lemere and Masliah, 2010; Wilcock et al., 2009) which attempt to immunise the brain against the formation of amyloid protein plaques. A major clinical trial also supports this view, showing that the effects of treatment with AChEIs such as *Donepezil* may actually have only very marginal benefits, and be relatively cost-ineffective as a form of treatment (Courtney et al., 2004).

In summary, whilst the cholinergic hypothesis is capable of explaining at least some of the cognitive deficit symptoms in Alzheimer's sufferers and the subsequent general symptomatic improvement following the administration of AChEI drugs, it does not adequately explain the underlying causes or pathology of the disease, nor the underlying reasons for

the greatly increased death of cholinergic neurons which cause the cognitive deficits.

2.1.3 Tau hypothesis

The tau hypothesis of Alzheimer's disease refers to the neurodegenerative effects of a modified (or *hyperphosphorylated)* form of the *tau* protein which aggregates with other fibres of tau and eventually creates the neurofibrillary tangles (NFTs) inside neurons which are prevalent in brains with AD (Spires-Jones et al., 2009). At present it is unclear whether the actual mechanism of tau-mediated neurodegeneration is mainly due to the "toxic gains-of-function acquired by the aggregates or their precursors [or] the detrimental effects that arise from the loss of the normal function(s) of tau in the disease state", or to a combination of both effects (Ballatore et al., 2007).

Tau protein is an essential component of structural *microtubules* within the neuron, a key building-block of cell axons during development, and important for transportation of neurotransmitter-containing vesicles along the axon in mature cells (Feinstein and Wilson, 2005; Johnson and Stoothoff, 2004). Instead of binding to the microtubules, tau in AD becomes sequestered into NFTs within the neurons (Ballatore et al., 2007) and as the level of normal tau in the brain is reduced the microtubules disintegrate, causing further neuronal dysfunction. As well as the loss of this normal supportive function during tau pathology, tau neurofibrillary tangles additionally present a physical obstacle to the transport of vesicles within the neuron (Ballatore et al., 2007), causing degeneration of syapses, axons and dendrites, and thus disrupted cell signalling.

Unlike the cholinergic hypothesis, which proposes that the major component of cognitive decline is due to changes in cholinergic signalling, the tau hypothesis proposes that the decline is due primarily to loss of synapses and neurons (via a toxic form of the modified tau), and the subsequent loss of connectivity which is experienced within the brain (Spires-Jones et al., 2009). Although the amyloid hypothesis (see section 2.1.4) provides a possibly more widely-encompassing view of AD, it has a number of significant unexplained problems, not least that "the number of amyloid deposits in the brain does not

correlate well with the degree of cognitive impairment" (Hardy and Selkoe, 2002), and so the tau hypothesis and its relationship to the amyloid hypothesis remain an important subject for further research.

2.1.4 Amyloid hypothesis

Beta-amyloid plaques

According to Boche and Nicoll (2010), "the principal hypothesis for the pathogenesis of Alzheimers disease for two decades or more has been built around amyloid, and known as the β-amyloid [Aβ] cascade hypothesis, which states that Aβ, either in the form of extracellular amyloid plaques or in soluble or oligomeric forms, has the key role in initiation of the disease". The hypothesis states that plaques consisting of extracellular deposits of Aβ derived from amyloid precursor protein (APP) are formed in the brain, leading to neuronal toxicity, cell death, and subsequent neurodegeneration and cognitive deficiencies (Francis et al., 1999).

There is wide-ranging support for the amyloid hypothesis, including genetic evidence relating to the risk factors of certain genes which lead to a build-up of excess amyloid. For example, the apolipoprotein E4 [apoE4] genotype, heavily linked with Alzmeiher's disease, is known to correlate with increased Aβ deposition (Hardy, 2009). Additionally, patients with Down's syndrome – caused by the duplication of chromosome 21, which is also the location of the gene for APP expression – almost universally develop early-onset AD (Hardy and Selkoe, 2002).

It is believed that amyloid pathology affects several aspects of the overall process of AD, altering neurological processes directly by blocking synapses (Hardy, 2009) and also exacerbating the effects of tau and cholinergic pathology in turn (Wilcock et al., 2009). This is known as the *amyloid cascade*: "The β-amyloid hypothesis describes a cascade of events [...] starting from APP [...] mutations as well as risk factors (e.g. apoE4) with the contribution of metabolic alterations, then the oligomerisation [aggregation of

protein monomers] of β-amyloid peptides followed by Tau phosphorylation. Both these modifications lead to the formation of intracellular neurofibrillary tangles and extracellular senile plaques as well as to synaptic and neuronal loss, resulting in network dysconnections and finally in the deterioration of episodic memory and dementia" (Savioz et al., 2009).

Calcium hypothesis of amyloid

However, original predictions that Aβ should in itself be directly toxic have proved to be hard to support (Hardy, 2009). One possible method, via the calcium signalling systems in the brain, is believed to result from the interaction of Aβ with neuronal membranes, leading to an influx of toxic Ca^{2+} ions into the neuron beyond a critical level at which *apoptosis* (programmed cell death) occurs (Hynd et al., 2004), or even *necrosis* (traumatic cell death) due to the high toxicity levels of Ca^{2+} (Berridge et al., 2000). Additionally, alterations in levels of Ca^{2+} are believed to accelerate the formation of Aβ, thus leading to a degenerative feed-forward cycle of toxic Aβ generation and Ca^{2+} disruption (Demuro et al., 2010).

However, Ca^{2+} is itself an important component of neural signalling mechanisms (Berridge et al., 2000), believed to play a direct role in cognition, including regulating neural responses to stimuli and affecting the action of glutamate receptors which are essential for long-term potentiation of synapses during learning in the hippocampus (Clapham, 2007), implying that a disruption in calcium signalling systems could lead to the learning and memory deficiencies observed in AD (Berridge, 2010).

Intracellular β-amyloid

As well as forming external amyloid plaques, there is evidence that Aβ accumulates internally within neurons, inhibiting *proteasomes* (protein complexes which regulate the concentration of proteins within a cell and clear up pathological proteins) and leading to a build-up of tau protein. Proteasome inhibition also leads in turn to a further build-up of Aβ, thus exacerbating the cascade, and to calcium dysfunction. It has also been suggested

that intracellular Aβ oligomers can adversely affect long-term potentiation (memory consolidation) in the hippocampus, leading to cognitive decline (LaFerla et al., 2007).

N-APP and DR6 interaction hypothesis

Despite the various mechanisms by which the amyloid cascade is proposed to affect cognition, the normal function of APP is still not actually known (Hardy, 2009). One proposed normal function which is hijacked by the pathology of Alzheimer's disease and which also leads to a form of amyloid toxicity is proposed by Nikolaev et al. (2009).

Typical cell death occurs via two processes: *necrosis*, whereby the cell dies of a traumatic interruption to its normal functioning (for example due to injury or infection), and *apoptosis*, whereby the cell is programmed to die either as part of a natural process of aging, or as a vital element of development. For example, in the early stages of brain development, apoptosis occurs "when productive neural circuits are reinforced and non-productive ones are eliminated. The neurons that fail to establish appropriate links with other neurons are culled mainly by apoptosis, a process that has the cell 'commit suicide' and then package up its residual bits for recycling" (Nicholson, 2009). This could be a parallel of the observed positive effects of pruning of the hidden layer of artificial feed-forward neural networks (Reed, 1993).

Nikolaev et al. (2009) present experimental data which suggests that a fragment of the amyloid precursor protein (APP), *N-APP*, is capable of binding to and activating the *DR6* (death receptor 6) receptors found on neuronal cell bodies and axons. DR6 is part of the mechanism for normal cell body apoptosis and axonal degeneration, and acts to accelerate cell degeneration when activated, as part of the brain's "pruning algorithms". The cell can also degenerate via apoptosis in other unrelated ways, but DR6 has the effect of accelerating this process as required by the cell. Part of the apoptopic mechanism involves the release of *caspases*, one of which – caspase 6 – is capable of cleaving the N-APP fragment from β-amyloid, thus leading to a cascade of activation of the DR6 receptor by N-APP, release of caspase 6 and cell death, and the cleavage of further N-APP from

existing β-amyloid deposits (Nicholson, 2009).

Synaptic scaling-driven progression

Small (2008) proposed a mechanism by which beta-amyloid pathology and the disruption of calcium signalling may be combined with existing synaptic compensatory mechanisms, leading to the directed spread of amyloid pathology. As neurons die, their remaining neighbours must scale up incoming activation by upregulating the accumulation of a type of glutamate neurotransmitter receptor known as AMPA (Turrigiano, 2008). This increase in sensitivity to activation as a response to nearby damage could lead to instability and over-activation of the compensating neurons, with resulting Aβ-mediated influx of Ca^{2+} leading to cell death via apoptosis.

2.1.5 Summary of pathologies and possible lesions

The major pathological processes predicted by each hypothesis of AD are summarised in table 2.1, alongside suggested equivalent computational lesions which could be applied to simulate each process in a neural network model.

Hypothesis	*Major processes*	*Computational lesions*
Cholinergic	Death of cholinergic neurons $\rightarrow$ reduced ACh $\rightarrow$ build-up of hyperphosphorylated tau and Aβ; reduced glutamate signalling $\rightarrow$ disrupted long-term potentiation (Francis et al., 1999)	Reduced input pattern clamping in learning $\rightarrow$ runaway synaptic modification (Hasselmo, 1994). Reduced time allowed for network to settle during recall process (Menschik and Finkel, 1999). Synapses less able to form Hebbian associations.

Tau	Neurofibrillary blocking of axonal transport (Ballatore et al., 2007). Collapse of microtubules supporting axonal transport (Feinstein and Wilson, 2005). Toxic tau → neuronal loss.	Muting / flattening, or randomly blocking, the sigmoidal function of a non-linear output unit. Random neuronal deletion.
β-amyloid	Toxic Aβ plaques → Ca^{2+} influx and neuronal death (Boche and Nicoll, 2010). Plaques directly blocking synaptic transmission (Hardy, 2009). Ca^{2+} signalling changes (Berridge, 2010) leading to reduction in neuronal spike firing (Berridge et al., 2000) and long-term potentiation (Clapham, 2007). Exacerbation of tau/cholinergic pathology (Savioz et al., 2009; Wilcock et al., 2009).	Synaptic / neuronal deletion with associated compensation (Ruppin and Reggia, 1995a). Randomly suppressing synapses during learning.
APP & DR6	Aβ cleaved to N-APP fragment → neuronal death via DR6 activation / caspase 6 → further cleavage of N-APP from Aβ (Nikolaev et al., 2009).	Pruning the hidden connections of a feed-forward neural network beyond a level at which the network is optimised, or pruning at a faster rate than compensatory mechanisms can cope with.

Scaling-driven Aβ	Death of neighbour neurons → up-regulation of glutamatergic AMPA receptors (Turrigiano, 2008) → instability of neural firing and increased Ca^{2+} influx → apoptosis (Small, 2008).	Allowing neurons to compensate for reduced input via a multiplicative "scaling factor" which restores activity but increases randomness of firing. Selecting neurons for deletion as a function of scaling factor.

Table 2.1: Hypotheses of Alzheimer's disease, their major mechanisms, and how they could be modelled computationally.

2.2 Artificial neural networks for modelling

This section reviews major types of artificial neural networks which may be used for computational modelling of Alzheimer's disease, and examines various biologically-inspired learning rules which may be applied in these networks, as well as some of the issues surrounding network connectivity and model abstraction.

2.2.1 Basic associative networks

Hopfield networks

Most artificial neural networks fall into one of two categories: *feed-forward* and *recurrent*. Feed-forward networks such as the multi-layer perceptron can be used for tasks such as nonlinear function regression and classification, but are not inherently biologically realistic due to their requirement for information to flow in strict layers with no feedback between each layer. Conversely, associative networks such as those of Hopfield (1982) contain numerous recurrent feedback loops and mimic the ability of biological neural networks to memorise and recall a given state from a given partial or noisy input cue; in an associative network, the on/off state of each neuron defines a part of the overall memory state of

the network, and the state of the network evolves over time as it settles into one of its basins of attraction (stable states into which the network will eventually settle if given the appropriate starting conditions).

The canonical associative neural network on which many computational models of brain processes are based was first described by Hopfield (1982). Briefly, a network of units, each consisting of a single uniform threshold value and taking a Boolean state $S_i = \{0, 1\}$, is fully connected with each connection holding a specific weight or strength. When the summed input to a unit is greater than the threshold, the unit transitions to the positive state (the neuron "fires"), thus affecting each of the other units to which it is connected. The dynamics of the Hopfield network state that, over time, the network updates fall into a basin of attraction and state transitions eventually cease. By a process of Hebbian learning (*"cells which fire together, wire together"*), the weights are set such that the network falls into a specific state vector (a list of the states of all the units) when a portion of that vector is presented to the network after learning, in a computational analogy of cognitive recall. (A full tutorial is given in Gurney (1997)).

Hopfield (1984) introduced the analogue Hopfield network in which each unit takes a continuous state value $\{0 \leq S_i \leq 1\}$ and the step threshold function is replaced with a sigmoid to smooth the state transitions, thus enabling the network to more closely resemble the behaviour of real biological neurons (Gurney, 1997).

Willshaw networks

However, Hopfield networks suffer from a number of flaws in the context of computational neuroscience, notably that they "are not plausible from the neurobiological point of view because they require symmetric weights, have only point attractors and are trained using non-local learning procedures" (Duch, 2007). In work by Horn et al. (1996), the gradient of memory loss during Alzheimer's disease is instead investigated in an associative Willshaw model (Willshaw et al., 1969). This model provides features to mitigate the biological implausibilites of Hopfield networks, such as asymmetric weights, leading to a model of

Alzheimer's disease with "two important differences [from the Hopfield model]: in the Willshaw model, the retrieval of some memories may decline while others are preserved. Furthermore, while in the [Hopfield] model once a memory pattern vanishes it is lost forever, in the Willshaw model memory patterns which are lost may later be adequately retrieved due to ongoing compensation".

The Willshaw model "consists of two layers of binary units. Units in the input layer make unidirectional connections with units in the output layer, with no recurrent connections within each layer" (Graham and Willshaw, 1995), implying one significant difference with the generic Hopfield network: the Willshaw network uses separate input and output pattern pairs, such that on presentation of a partial or noisy input to one part of the network, the corresponding output pattern is retrieved elsewhere in the network. Conversely, in the Hopfield network, the output state overwrites the noisy input state vector.

The Willshaw network permits the use of realistic variable neural firing thresholds but at the expense of requiring biologically implausible binary connections. It has been updated to support the ability to degrade gracefully when storage capacity is reached, as would be expected in Alzheimer's disease, but for optimum performance the network must know parameters such as the number of neurons and the firing ratio (coding rate) which would be impossible to know on a global scale in a biological neural network (Henson and Willshaw, 1995). Additionally, Golomb et al. (1990) highlight a key flaw of the learning rule used in the Willshaw model, in comparison with the activity-dependent learning rule used in the Hopfield-based Tsodyks and Feigel'Man (1988) model: "the Hebb rule adopted by Tsodyks and Feigel'Man induces also *suppression* of the synaptic efficacies, whereas Willshaw's rule uses only enhancement of synaptic efficacies. In biology, both long-term potentiation (LTP) of synapses and long-term depotentiation (LTD) have been observed. The superior performance of the model of Tsodyks and Feigel'Man suggests that both mechanisms are essential for an effective learning".

With that in mind, we now examine a selection of biologically-inspired learning rules.

2.2.2 Biologically-inspired learning rules

Hebbian learning

Hebb (1949) proposed the well-known learning mechanism for synaptic modification, which states that:

> "when an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased".

This can be neatly summarised in the premise *"cells which fire together, wire together"*. This principle manifests itself in the learning rule in equation 2.1, where W_{ij} is the weight of the synapse between neurons i and j, γ is a learning rate, and x^p is the Boolean representation of the pth unit of an input pattern x. Here, the change in weight is only ever positive, and occurs only when both neurons fire simultaneously:

$$\Delta W_{ij} = \gamma x_i^p x_j^p \tag{2.1}$$

Activity-dependent Hebbian learning

Tsodyks and Feigel'Man (1988) and Ruppin and Reggia (1995a) introduce an activity-dependent learning rule for pattern storage, based on the Hebbian principle but introducing the requirement for each given pair of units to remain in the same state for a certain number of update iterations before the synaptic weight between them is modified. This requires each pattern to be presented several times in turn to the network before it is stored. Additionally, unlike the Hebb rule, this activity-dependent learning rule is capable of depressing synapses between pairs of neurons which do not co-incidentally fire:

$$\Delta W_{ij} = \frac{\gamma}{N}(S_i - p)(S_j - p) \tag{2.2}$$

where γ is the learning rate, N is the number of neurons in the network, S refers to the

neuronal state $\{0, 1\}$ (or "firing" and "not firing"), and p is the coding rate denoting the average proportion of *1*s compared to *0*s in the stored memory patterns.

Thus the learning algorithm attempts to mitigate the effects of the Hebb rule's ability to globally alter synaptic weights in a biologically-unrealistic way and avoids using its method of storing each pattern in a 'one-shot' process which is susceptible to the presence of errors or noise. By presenting each pattern several times to the network, any noise present in the inputs is reduced and the synaptic matrix is gradually constructed rather than being enforced in a single process by the learning rule. This learning rule is presented in more detail in the next chapter (equation 3.1).

Oja's learning rule

Synaptic weights undergoing Hebb rule updates can grow to infinity during learning, as increases in correlated firing between neurons lead to strengthening of synapses, and thus to further increases in firing correlation. Oja (1982) therefore introduced an additional 'forgetting' term to the Hebb rule to balance out this infinite growth. Equation 2.3 is formulated for a single neuron, where y is the output of the neuron, calculated as the weighted sum of all the inputs x_i:

$$\Delta W_i = \gamma(x_i y - y^2 W_i) \tag{2.3}$$

The negative term means that the greater the output of the neuron becomes, the greater the balancing reduction in the weight becomes. This leads to competitive synaptic updates at each neuron, such that an increase in the weight of one synapse causes a corresponding decrease in the weights of the other synapses.

To make Oja's rule applicable to networks of neurons, the learning rule is altered as follows (equation 2.4), where $k \neq \{i, j\}$ is the index of each synapse which is depotentiated

to balance the potentiation of W_{ij} (Sanger, 1989):

$$\Delta W_{ij} = \gamma y_i (x_j - \sum_{k=1}^{i-1} y_k w_{kj}) \tag{2.4}$$

Therefore, any increase in the synaptic weight from neuron i to neuron j leads to a corresponding decrease in the weights of the other neurons which act on j. Although this rule allows networks of neurons to maintain normalised synaptic weights via competition, there is a disadvantage: learning in the network is sequential, meaning that each cell cannot be updated until all the preceding cells have themselves been updated (due to the sum over k up to i), making biologcally-realistic synchronous weight updates impossible.

Similar effects can be achieved using the BCM neuron model (Bienenstock et al., 1982), in which the neural firing threshold (rather than synaptic weights) are modified to maintain average presynaptic activity.

LEABRA

Whilst learning rules based on the Hebbian principle are nearly ubiquitous in the training of associative networks, Hebbian learning is self-organising and makes no use of error-correction mechanisms (for example backpropagation). O'Reilly (1996) argues that error-driven learning algorithms are essential for addressing the complexity of the problems solved in the human neocortex due to their much greater computational power. For example, networks trained with a Hebbian rule are unable to cope with higher-order input/output mappings (Churchland and Sejnowski, 1994), and can only solve problems of correlation (i.e. "does feature A correlate with feature B?") due to the lack of hidden units for higher-dimensional projection. However algorithms which might enable such computation, such as *contrastive Hebbian learning* (Movellan, 1990), are hard to justify from a biological perspective. Associative networks using error-driven learning are also poor generalisers to unseen input (O'Reilly, 2001). The *LEABRA* (local, error-driven and associative biologically realistic algorithm) learning rule introduced by O'Reilly (1996)

attempts to solve these issues by merging error-driven learning and Hebbian self-organising learning into a single homogenised algorithm whilst remaining biologically plausible.

O'Reilly et al. (1999) present a model of working memory – the part of the brain used for task-processing and assimilation of information prior to learning, providing "temporary retention of information that was just experienced but no longer exists in the external environment, or was just retrieved from long-term memory" (D'Esposito, 2007) – in a modular system incorporating prefrontal cortex and hippocampal sections, implemented using LEABRA. Whilst not a particularly modern model compared to those reviewed by D'Esposito (2007), it provides a good example of the use of LEABRA for creating biologically-plausible neural models.

The implementation details for LEABRA are fairly complex, but the algorithm is described in O'Reilly (2001) and further explored in O'Reilly and Munakata (2000). The *Emergent*[1] simulation engine implements LEABRA as well as a wider variety of neural network simulation methods.

2.2.3 Sparse connectivity strategies

Biological neural networks such as those found in the hippocampus and neocortex are generally sparsely connected (Levy, 1996; Bullmore and Sporns, 2009). Evans (1989) shows that a process of diluting the synaptic weight matrix of a Hopfield network such that it is no longer fully-connected still causes it to behave in much the same way as a fully-connected network when stored state vectors are generally of low activity, "implying that random dilution has not changed the features of the model to a great extent". Dilution mechanisms for associative networks include connecting each unit in a spatial manner to those immediately surrounding it with Gaussian probability as performed in the Tsodyks and Feigel'Man network by Ruppin and Reggia (1995b), or by using a randomised small-world network connection strategy (Watts and Strogatz, 1998).

[1] `http://grey.colorado.edu/emergent/index.php/Main_Page`

Graham and Willshaw (1995) present a non-fully-connected Willshaw network in a biological context, although there are certain problematic informational requirements from the perspective of each neuron, and the authors state several certain neurobiological constraints which must be met if their model is to be considered plausible. In a later paper, Graham and Willshaw (1997) conclude that the number of units, rather than the specific connectivity, is the main determiner of capacity.

2.2.4 Networks of spiking neurons

The simplicity of the Hopfield-type network provides one of its great advantages in computational modelling, as it allows the study of qualitative network behaviour during tasks such as associative recall, with greatly reduced complexity. However, some of the necessary assumptions of a Hopfield-type network are highly artificial, and drastically change the dynamics of the system compared to more realistic networks such as those found in the brain.

Such assumptions include the binary on–off response of neurons to patterns of input, which is a gross simplification compared to biological neurons (Gerstner and van Hemmen, 1992). Biologically-realistic neurons employ a *spike-firing* mechanism, rather than simply switching on or off depending on the result of the evaluation of a summed input against a threshold. Spiking neural networks incorporate this additional complexity by transmitting a series of pulses, or *spikes*, of activation when excited. The intervals between spikes (rather than the explicit binary presence or absence of a spike, or the magnitude of the spike) thus determines the nature of the output activation, and allows a much greater richness in the type of processing which can be achieved by the network (Ghosh-Dastidar and Adeli, 2009).

Spiking neural models necessarily introduce a large amount of extra model complexity, and the behaviour of simplified Hopfield-type models can be viewed as qualitatively similar to the behaviour observed in spiking network models (Duch, 2007), so a standard thresholding associative network could be used to meet many requirements for a

computational model of neurological disorders such as AD.

Gerstner and van Hemmen (1992) agree that in some states (particularly stationary pattern-retrieval states), the extra dynamics included by spiking neurons may be disregarded, and the system may be viewed simply as a network of binary neurons which respond (or not) in some way to the input pattern. But in other states, particularly those involving brain-like oscillations of activity within a spiking network, the timing between spikes and the noise in the system – as determined by constants such as refractory periods between spikes and membrane depotentiation amounts – play a very large part in determining the behaviour of the system. This indicates that the benefits from employing a spiking behaviour are observed in some cases, but not all. However, in those cases where behaviour of a spiking model differs from that of a simplified associative model, spiking models could provide useful extra insight for further research into cognitive disorders.

Gerstner and Kistler (2002) provide an in-depth tutorial on networks of spiking neurons, and Ghosh-Dastidar and Adeli (2009) give a concise overview of spike-generation mechanisms. The increased biological realism offered by such networks could be a useful incorporation into existing and future models of Alzheimer's disease. van Rossum et al. (2000) show how Hebbian learning can be implemented in a model of synaptic plasticity which is *spike time-dependent* such that potentiation and depotentiation occur with the relative timings of spikes between presynaptic and postsynaptic neurons, which is related to observations in the brain (Markram et al., 1997).

2.2.5 Reservoir computing

Schiller and Steil (2005) showed that it is possible to functionally partition a recurrent neural network into two layers: one containing a dynamical reservoir of randomly-connected untrained recurrent neurons (which can be of any type including spiking neurons or simple linear or sigmoidal threshold logic units (Schrauwen et al., 2007)), and another output layer implemented as a trained linear read-out of the reservoir (figure 2.1). This allows input into the reservoir to be transformed into high-dimensional space (including complex

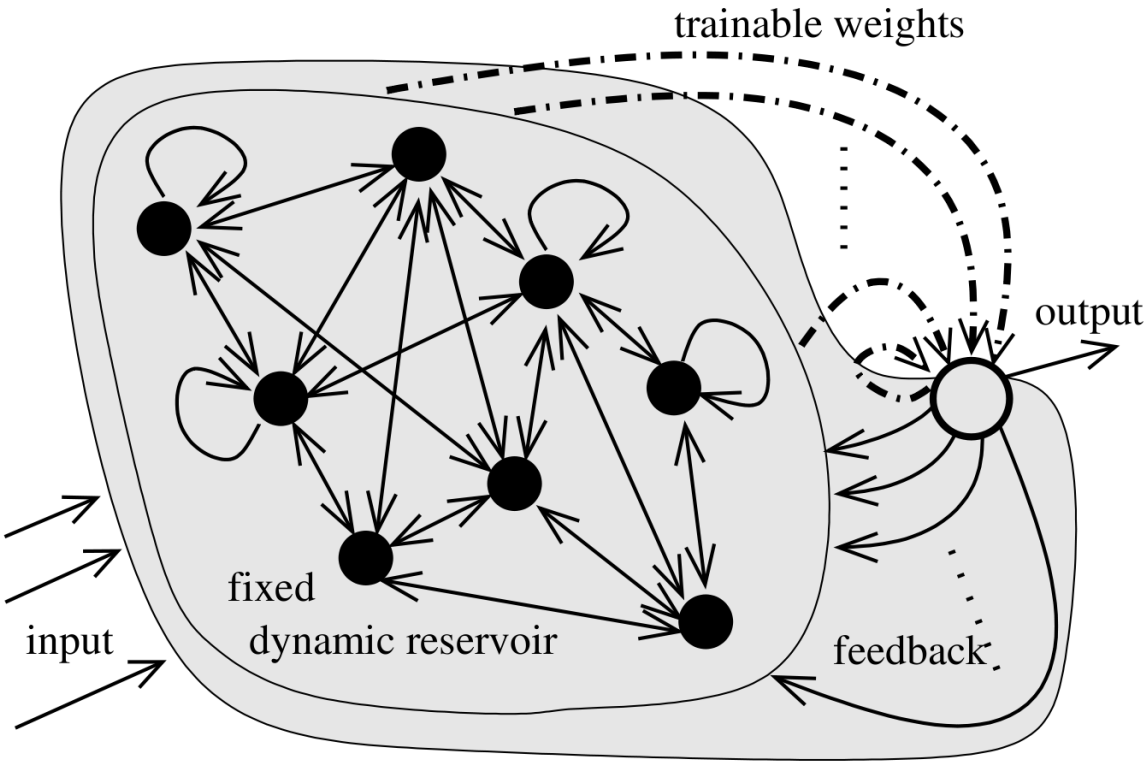


Figure 2.1: A reservoir of randomly-connected untrained recurrent neurons projects the inputs into high-dimensional space, from which the trained linear read-out layer learns the mapping from input to output, with optional feedback to the reservoir to enable error-correction as a part of a backpropagation strategy. Diagram by Steil (2004).

transformations such as temporal integration), providing a non-linear feature detector with enhanced separability, from which the read-out layer can learn (Morse and Aktius, 2009; Verstraeten et al., 2007).

Models utilising this principle are known by the term *reservoir computing*. Two main implementation classes of reservoir models exist: *echo state networks* (ESN) and *liquid state machines* (LSM) were both developed simultaneously by different groups and share a number of similar features (Jaeger, 2001; Maass et al., 2002). The differences appear to be mainly superficial, and relate to specific implementation methods. Whilst ESNs are typically applied to engineering problems such as robotic navigation (Reinhart and Steil, 2009), LSMs were developed from a computational neuroscience perspective, meaning that such models often make use of spiking neurons and synaptic plasticity within the reservoir (Lukosevicius and Jaeger, 2009). One such example of research aiming to

utilise the biologically inspired features of LSMs is given by Yamazaki and Tanaka (2007), who propose a model of the cerebellum as a liquid state machine, simulating long term depotentiation of neural circuits with the inclusion of temporal information, by way of comparison with a perceptron model which stores only temporally non-specific mappings.

Lukosevicius and Jaeger (2009) give a comprehensive overview of various reservoir computing and related methods, including strategies for reservoir creation and various methodologies for constructing and training the readout layer, with the aim of unifying their various concepts under the single banner of *reservoir computing*. Verstraeten et al. (2007) additionally provide experimental comparisons of the performance of various reservoir computing methods, whilst Schrauwen et al. (2007) offer a tutorial on certain issues of implementation.

A specific training method for reservoir-based networks using online learning, introduced by Steil (2004) as *backpropagation decorrelation* (BPDC) and analysed in Steil (2007), is suited to adaptation to signals which change over time. Reinhart and Steil (2009) apply ESN methods using the BPDC rule in an engineering context to create a sensory-motor controller for a robotic arm, utilising feedback from the linear output layer back to the reservoir and showing that this permits a form of error-correction.

Applications such as these indicate that a reservoir computing model, when used for the modelling of Alzheimer's disease, could provide a significant advantage over the use of a traditional Hopfield-style associative network (for example Ruppin and Reggia (1995a)) as not only is the reservoir computing model more biologically resemblant of cortical modules in the brain than a basic Hopfield attractor (Lukosevicius and Jaeger, 2009), but a model utilising reservoir techniques will be able to perform more complex learning tasks such as sensory-motor control or learning of grammatical construction in natural language (Dominey et al., 2006). Benefits of modelling neurological disorders during performance of specific psychological tasks in this way, as opposed to simple associative recall, have previously been shown by Hoffman and McGlashan (2001) in the context of schizophrenia.

A further benefit of modelling neurological lesions in a reservoir model compared

with a standard Hopfield-type model is due to the nature of its construction. Consider a Hopfield-class model such as Ruppin and Reggia (1995a), based on a Tsodyks and Feigel'Man (1988) network, which contains an arbitrary number of processing units, N. This type of model uses an activity-dependent learning rule (section 2.2.2) to learn and recall patterns from a noisy input cue. A percentage of the neurons in this model are deleted as part of a process of lesioning, and at a certain limit catastrophic degradation occurs. Results showed that this limit on deletion can be as large as $0.5N$. By the simple nature of the fact that every unit of the network is both a processing unit as well as an input/output unit, even if the deletion limit has not yet been reached up to half of the output units of the network are now missing, resulting in a vastly altered output pattern which is unrepresentative of the underlying pathology (in which the network is still functioning correctly, as the deletion limit has not yet been reached). This is due to the simplicity of the assumption that the Hopfield network makes regarding the role of each unit as used for both processing and output. Conversely, in a reservoir model with N units, of which perhaps only $0.05N$ are in the designated output layer, the chances of the output patterns being seriously altered in this way during lesioning are much smaller, simply due to the far smaller ratio of output-to-processing units ($1 : 20$) in this model compared with the Tsodyks and Feigel'Man (1988) model ($1 : 1$).

2.2.6 Abstraction in computational modelling

Simple connectionist associative neural networks such as the Hopfield and Willshaw models belong to a different category from biophysically-realistic spiking neural network models, offering *biological approximation* rather than *biological realism*. This abstraction of complexity, whilst limiting the resolution of the model on the one hand, can provide insights of a different kind to neuroscientists by revealing larger-scale network effects which may otherwise be lost in the detail of fine-grained models. Care must be taken to ensure that, whilst approximations of the "integrate and fire" sort which retain the key biological aspects of neural behaviour may be permitted (Abbott, 1999), principles of large-scale

biological plausibility such as sparse coding in stored patterns (Abeles et al., 1990) and sparse connectivity between neurons (Churchland and Sejnowski, 1994) must be adhered to.

By necessity, a number of assumptions must be made when using such a model, but the whole reason for modelling is to simplify a complex system and reduce it to a level at which it can be understood. As Anderson and Kreiman (2011) succinctly put it:

> "What we cannot model, we do not understand".

Providing such assumptions are made with foresight and justification, such basic models can actually become very powerful tools for understanding principles of computation in neural substrate. To paraphrase a saying:

> "If you know what you're doing, and you simplify it, that's abstraction. If you don't know what you're doing, and you simplify it, then you don't know what you're doing!"

The aim of this work therefore is not to replicate directly the biological processes implicated in Alzheimer's disease, but rather to abstract out the effects of proposed principles in a clearer way than would initially be possible in neurobiological models. This thesis begins in the next two chapters with experiments in a basic associative model, due to the simplicity of these types of networks. In later parts of this thesis we examine these principles in more complex and realistic spiking neural network models, with the aim of making biologically-testable predictions and revealing potential changes in high-level behaviour. These observations could then be used to predict the presence of Alzheimer's disease in patients before the onset of cognitive symptoms, thus enabling earlier intervention and treatment.

2.3 Chapter summary

This chapter has introduced the major hypotheses of Alzheimer's disease, along with potential computational representations of the relevant lesions. A selection of associative

artificial neural networks and biologically-inspired learning rules from the literature has been presented, including a focus on sparse connectivity strategies, spiking neural networks, and the potential of reservoir computing techniques to compensate for some of the shortfalls of simple Hopfield networks. Finally, we end with a closing note on the importance of good abstraction in computational modelling.

CHAPTER 3

EXPLORING A SIMPLE ASSOCIATIVE MODEL

Chapter synopsis

In this chapter, we will examine a basic model from the Hopfield class, developed by Tsodyks and Feigel'Man (1988), to observe the effects of Alzheimer's-related damage on networks with various different connectivity and compensatory strategies.

We begin with a replication of the results of Ruppin and Reggia (1995a), and present a clearer description of an existing local neural field-dependent synaptic compensation rule than is available in the literature. We examine the effects of using recent versus remote memories during compensation, leading to suggested therapeutic strategies which may reduce the progression of damage. We also investigate the effects of connectivity density and type (e.g. small-world) on network capacity and robustness to damage. Finally, we present a possible lesioning method representing tau pathology.

Work derived from this chapter has previously been published as Rowan (2011).

3.1 Introduction

As previously described, Alzheimer's disease (AD) is a specific form of dementia, characterised biologically by neurofibrillary tau protein tangles and beta-amyloid (Aβ) protein plaques (Tiraboschi et al., 2004), and symptomatically by a progressive decline in memory

capabilities. In particular, recent memories are the first to be lost whilst distant memories are retained, but as the disease progresses this is followed by gradual total loss of recall, a corresponding loss of personality, motor control, and other bodily functions, and finally death (Francis et al., 1999).

Computational modelling of neurological disorders such as AD is an established tool (Aakerlund and Hemmingsen, 1998), using models ranging from simple Hopfield-style recurrent neural networks, to fully-detailed biologicaly-validated spiking models of whole brain regions including multiple neurotransmitter types and ion channels. Many such models can be found on the Yale ModelDB[1].

One widely-cited Hopfield-class model, by Ruppin and Reggia (1995a), showed how simple lesions in a single-layer associative network trained in an *activity-dependent* Hebbian manner leads to loss of stored memory, and the addition of a local compensation factor causes the pattern of functional damage to mirror more closely that found in AD whereby recently-stored memories are lost before historical memories. Later work showed how the compensation factor can be made biologically plausible by depending only on the individual post-synaptic potential of the remaining neurons after lesioning (Horn et al., 1996), rather than providing each neuron with some global knowledge of the amount of damage.

Whilst experimental support exists for the predicted compensatory strengthening of synapses in AD (Savioz et al., 2009; Turrigiano, 2008), one unexplained result drawn from this model is that significant neuronal deletion (around 50%) in the model is required before memory function is seriously impaired. This rate of deletion is much larger than the rate observed clinically in the latest stages of Alzheimer's disease (between 10% and 30% reduction of volume in the hippocampal regions in severe cases of AD (Minati et al., 2009) and certainly far more than the 10% general cerebral atrophy reported at initial diagnosis of the disease (Ridha et al., 2006)), implying that there must be other factors additionally affecting cognitive decline.

[1] https://senselab.med.yale.edu/modeldb

A further limitation of the model is that lesioning is performed only by deleting a number of randomly-selected neurons or connections at each step, which does not necessarily represent the subtleties of the underlying pathology. The authors present a method of applying lesions in a localised spatial manner by deleting all of the neurons and/or connections within a circle or rectangle of a given area (Ruppin and Reggia, 1995b), but this does not incorporate any of the known neurodegenerative mechanisms such as tau or amyloid pathology.

Additionally, the method in which the network is interconnected (either fully, or using an arbitrary number of connections per neuron in a localised Gaussian manner) is again simplistic and does not represent biologically realistic connection strategies such as small-world networks (Watts and Strogatz, 1998) or neural Darwinism (pruning of weaker synapses during development) (Hoffman and McGlashan, 2001).

This model continues to be widely cited (Duch, 2007; Small, 2008; Savioz et al., 2009) even though it could be made capable of better approximation of the lesions representing AD pathology. Today we know much more about connectivity strategies in the brain such as small-world networks (Watts and Strogatz, 1998; Bullmore and Sporns, 2009) as well as the biological processes underpinning AD, such as neurofibrillary tau pathology.

Here, methods are presented for enhancing the Ruppin and Reggia (1995a) model with details which may be more representative of the underlying biology. The intention is to examine differences in behaviour which may occur when considering connectivity strategies, specific details of compensatory techniques, and lesioning in accordance with specific pathologies, with the aim of leading to development of more accurate representations of a range of pathological processes underlying AD.

The remainder of this chapter is organised as follows: Section 3.2 describes the Ruppin and Reggia model in greater detail and the updates made to it in this work; section 3.3 presents the results of experiments characterising the network's behaviour with these new enhancements; and section 3.4 provides concluding remarks and outlines future directions in which this study could be taken.

3.2 Model description

3.2.1 Learning rule

Ruppin and Reggia (1995a) previously showed how a variant of an attractor network model proposed by Tsodyks and Feigel'Man (1988) (the T-F model) is capable of storing patterns in a biologically-plausible Hebbian *activity-dependent* manner. This is achieved using a repetitive-learning process whereby each pattern to be stored "must be presented to the network several times before it becomes engraved on the synaptic matrix with sufficient strength, and is not simply enforced on the network in a 'one-shot' learning process" (Ruppin and Reggia, 1995a). An updated version of the model (Ruppin and Reggia, 1995b) added Gaussian partial-connection of the network rather than full connectivity.

The model can be seen as a rough approximation of a single cortical unit. A set of external inputs, representing input from other cortical units or from the CA3 region of the hippocampus (Rolls, 2010; Rolls and Kesner, 2006; O'Reilly and Rudy, 2000), delivers activation greater than the neural threshold to each unit of the network according to the pattern to be learned.

The network learns patterns through a process of activity-dependent learning according to the update rule:

$$W_{ij}(t) = W_{ij}(t-1) + \frac{\gamma}{N}(S_i - p)(S_j - p) \tag{3.1}$$

where W is the weight matrix of undirected connections between neurons i and j, γ is a constant determining the magnitude of activity-dependent changes (learning rate), N is the number of neurons in the network, $K \leq N$ is the fixed number of other units to which each unit is connected, S refers to the neuronal state $\{0, 1\}$ (or "firing" and "not firing"), and p is the coding rate denoting the average proportion of *1*s compared to *0*s in the stored memory patterns. This is equivalent to flipping a coin independently for each element of the pattern and setting it to "1" with probability p, with $p \ll 1$ as cortical

networks are found to have low coding rates (Abeles et al., 1990).

The activity-dependent learning rule for pattern storage is based on the Hebbian principle of potentiation when two functionally connected cells fire together. However, unlike in a standard Hopfield network, the learning rule introduces the requirement for each given pair of units to remain in the same state for a certain number of update iterations (the suggested value is 5) before the synaptic weight between them is updated, and requires each pattern to be presented several times in turn to the network before it is completely stored (figure 3.1).

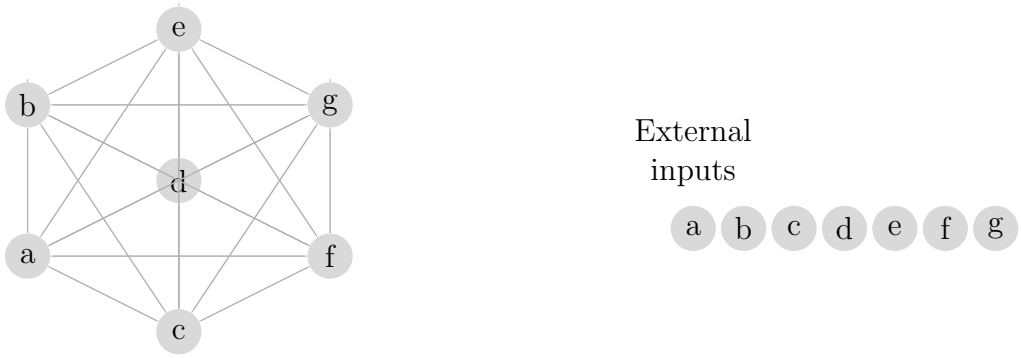
Thus the learning algorithm circumvents the Hopfield network's method of storing each pattern in a 'one-shot' process which is susceptible to the presence of errors or noise. By presenting each pattern several times to the network, any noise present in the inputs is reduced and the synaptic matrix is gradually constructed rather than being enforced in a single process by the learning rule. Additionally, unlike the Hopfield learning rule, the use of the $(S_i - p)(S_j - p)$ term to update the weights enables depotentiation in the case that either or both of the neurons i and j are quiescent ("0"), for any non-zero value of p.

The learning rule forms a *basin of attraction* for each pattern; a stable state into which the network will eventually settle if given the appropriate starting conditions.

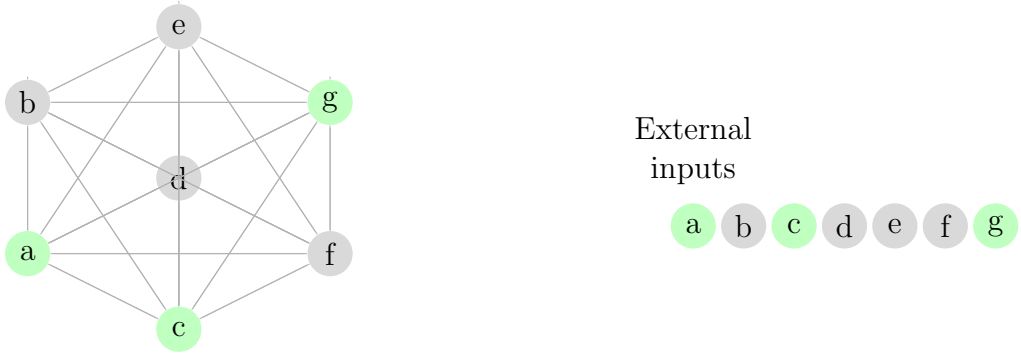
3.2.2 Performance evaluation

Patterns are recalled using a noisy version of the complete pattern applied to the network via the same set of external inputs used for learning with activation less than the neural firing threshold (figure 3.2). A similarity measure given by Ruppin and Reggia (1995a) for the recall performance of the network on a given pattern ξ^μ, termed the *overlap* between the resulting network state and the pattern, is as follows:

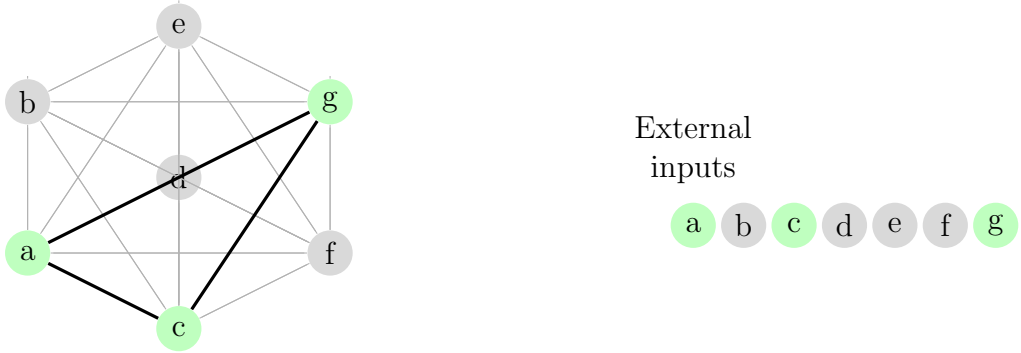
$$m^\mu(t) = \frac{1}{p(1-p)N} \sum_{i=1}^{N} (\xi_i^\mu - p) S_i(t) \tag{3.2}$$



(a) The network is initialised to a random state (here, all neurons are quiescent).



(b) A binary pattern is presented to the network via the setting of external inputs above the neurons' firing thresholds, causing the neurons to fire.



(c) For neurons which co-incidentally fire for at least five update iterations, their connecting weights are strengthened.

Figure 3.1: Learning in a schematic of the Tsodyks and Feigel'Man (1988) network using the activity-dependent learning rule. Here, the network is fully-connected, but the network can take various random sparse connectivity strategies.

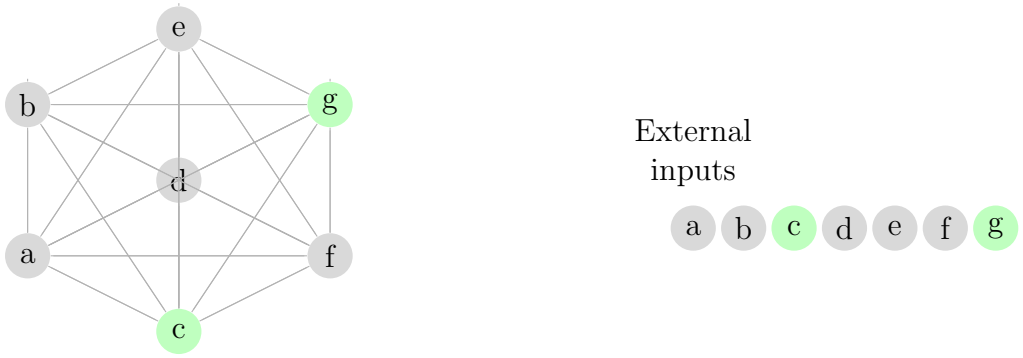
This measure has the effect of counting the correctly-firing units whilst also penalising with a lower weighting those units which fire erroneously (Tsodyks and Feigel'Man, 1988). The normalisation by $p(1-p)N$ ensures that, for example, the network cannot obtain a high overlap score simply by setting every unit to "0" (which would, of course, match 90% of the units on average if $p = 0.1$, without such an adjustment), whilst allowing the overlap score to range smoothly from $0 \rightarrow 1$ as the network more accurately recalls patterns.

3.2.3 Compensatory synaptic scaling

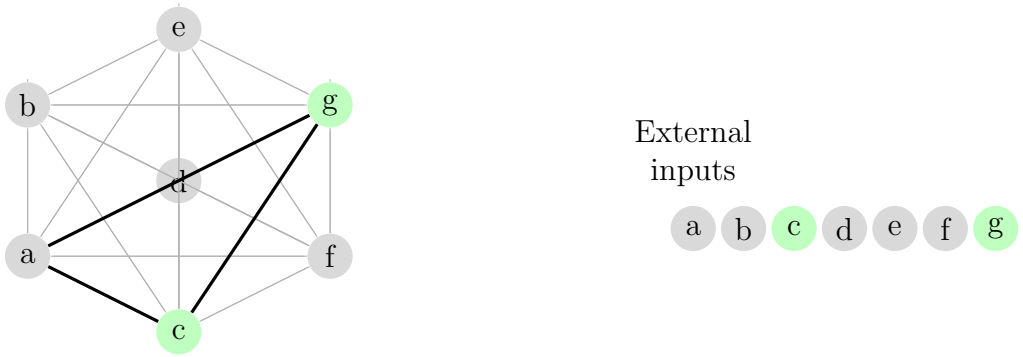
In the work by Ruppin and Reggia (1995b), the network model was lesioned by deleting synapses or neurons at random and implementing a process of *variable synaptic compensation*, where "the magnitude of the remaining synapses is uniformly strengthened in a manner that partially compensates for the decrease in the neuron's input field" (Ruppin and Reggia, 1995a) by multiplying the weights of the remaining synaptic connections by a globally-determined (i.e. depending on knowledge of the overall fraction of deletion) local compensation factor.

Ruppin and Reggia (1995a) examined the overall degradation in recall performance and the pattern of relative sparing of older memories compared to recently stored patterns (as observed in AD patients (Kopelman, 1989)) as the network was progressively lesioned, and concluded that synaptic deletion and compensation in this model can be demonstrated to reveal similar symptoms to the cognitive decline observed in AD.

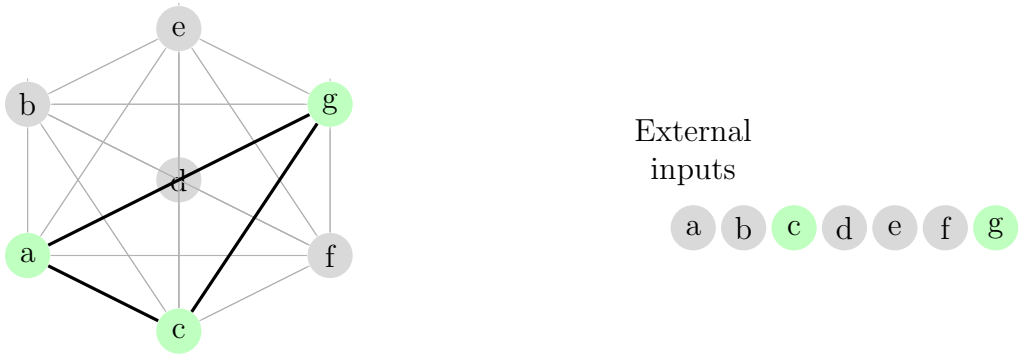
However, a global compensatory synaptic scaling strategy is biologically implausible as each neuron must somehow be aware of the global deletion rate both for itself, and for other neurons around it. Horn et al. (1996) therefore introduced a neuronal-level compensatory mechanism which causes each neuron to adjust its output based only on changes in the neuron's average post-synaptic potential (or summed input), and which does not require the explicit knowledge of either global or local levels of synaptic deletion.



(a) A *partial* binary pattern is presented to the network via the setting of external inputs above the neurons' firing thresholds, causing the neurons to fire.



(b) The stored pattern has a basin of attraction, defined by the learned synaptic weights. This causes activation to spread to other neurons which also form part of the same basin of attraction.



(c) The network updates until a stable state (no changes since the last iteration) is reached. If the network is performing correctly, the original pattern should have been retrieved.

Figure 3.2: Retrieving patterns in the Tsodyks and Feigel'Man (1988) network.

Algorithm

Synaptic compensation can be achieved by locally estimating, for each neuron, its remaining proportion of connectivity $0 \leq \hat{w}_i \leq 1$ at a given time, based on changes in postsynaptic (incoming) activity, and setting a "compensation factor" (a multiplier value of postsynaptic activity) $c_i = \frac{1}{\hat{w}_i}$ to account for the reduction in estimated connectivity.

The estimation of the remaining connectivity for each neuron i can be achieved by the measurement of the neuron's average squared summed weighted input $h_i^2(\hat{w}_i)$ over a set of pattern retrieval trials, comparing this over time with the previous value of h_i^2 in which $w_i = 1$ (i.e. when the network suffered no damage), to estimate $\hat{w}_i$ (Horn et al., 1996):

$$h'^2_i(\hat{w}_i) = c_i^2 \hat{w}_i h_i^2(w_i = 1) \tag{3.3}$$

The presence of noise in the network is accounted for by separating the process of pattern presentation into sets of "signal" and "noise" presentation. We can obtain $h_i^2(\hat{w}_i)$ in the "signal" state by retrieving the set of previously-stored patterns in the network[1], with the "noise" state value obtained by the presentation of randomly-generated patterns (which were not previously stored in the network, and which satisfy the coding rate $p \ll 1$), and allowing the network dynamics to update from that fixed point.

The expected value of $h_i^2(\hat{w}_i)$ over all patterns can therefore be obtained under the two conditions $\langle h_i^2(\hat{w}_i)\rangle_{S,N}$ and $\langle h_i^2(\hat{w}_i)\rangle_N$, where S, N is the "signal" state with intrinsic noise, and N is the "noise-only" state, with different power dependence on deletion (Horn et al., 1996):

$$\langle h_i^2(\hat{w}_i)\rangle_{S,N} = c_i^2 \hat{w}_i^2 \langle h_i^2(\hat{w}_i)\rangle_S + c_i^2 \hat{w}_i \langle h_i^2(\hat{w}_i)\rangle_N \tag{3.4}$$

The expected value of $\langle h_i^2(w_i)\rangle_S$ is derived from equation 3.4:

$$c_i^2 \hat{w}_i^2 \langle h_i^2(\hat{w}_i)\rangle_S = \langle h_i^2(\hat{w}_i)\rangle_{S,N} - c_i^2 \hat{w}_i \langle h_i^2(\hat{w}_i)\rangle_N \tag{3.5}$$

[1] Horn et al. (1996) speculate that this process could occur biologically during dreaming.

After this point, the network is damaged by deleting neurons or synapses. The new estimate of connectivity, $\hat{w}_i$ must now be made for each neuron. To achieve this, new updated values for $\langle h'^2_i(w_i)\rangle_{S,N}$ are obtained by repeated presentation of patterns to the network. Using the previous pre-damage values of $\langle h_i^2(\hat{w}_i)\rangle_S$ and $\langle h_i^2(\hat{w}_i)\rangle_N$, it is then possible to solve equation 3.5 for $\hat{w}_i$. Dividing by $c_i^2\hat{w}_i\langle h_i^2(\hat{w}_i)\rangle_S$ gives:

$$\hat{w}_i = \frac{\langle h'^2_i(\hat{w}_i)\rangle_{S,N}}{c_i^2\hat{w}_i\langle h_i^2(\hat{w}_i)\rangle_S} - \frac{\langle h_i^2(\hat{w}_i)\rangle_N}{\langle h_i^2(\hat{w}_i)\rangle_S} \tag{3.6}$$

but with $c_i^2\hat{w}_i = 1$ (equation 3.3), this becomes:

$$\hat{w}_i = \frac{\langle h'^2_i(\hat{w}_i)\rangle_{S,N} - \langle h_i^2(\hat{w}_i)\rangle_N}{\langle h_i^2(\hat{w}_i)\rangle_S} \tag{3.7}$$

In this way, the change in the "signal-only" term $\langle h_i^2(\hat{w}_i)\rangle_S$ is compared over time. From $\hat{w}_i$ we derive the new value for the compensation factor:

$$c_i = \frac{1}{\hat{w}_i} \tag{3.8}$$

3.2.4 Model dynamics

The update rule for neuron i at time t without synaptic compensation is defined as:

$$S_i(t) = G\left(\sum_j [W_{ij}S_j(t-1)] + e_i - \theta\right) \tag{3.9}$$

where S denotes the state of neuron i at time t, W_{ij} is the synaptic weight between neurons i and j, e_i is the external input strength (if activated), and θ is the neural threshold. $G(x)$ is the sigmoid function, defined as:

$$\frac{1}{1+e^{-\frac{x}{T}}} \tag{3.10}$$

where x is the weighted summed input to the neuron, and T is a "noise" value, which determines the slope of the sigmoid.

With the addition of local synaptic compensation, the update dynamics become:

$$S_i(t) = G\left(\sum_{j \in D_i} [c_i W_{ij} S_j(t-1)] + e_i - \theta\right) \tag{3.11}$$

where D_i denotes the set of neurons to which neuron i remains connected after lesioning. A more detailed description of the dynamics of the T-F model (albeit using the step function rather than the sigmoid function) is given in Horn et al. (1993).

3.2.5 Connectivity strategies

Biological neural networks such as those found in the hippocampus are generally sparsely connected (Levy, 1996; Bullmore and Sporns, 2009). It has been shown that in Alzheimer's disease, small-world clustering (as measured by the clustering coefficient, equation 3.12) is significantly reduced at a global level, resulting in large changes to the local organisation of the network (Supekar et al., 2008).

Connection dilution mechanisms for associative networks include connecting each unit over a flat random distribution, connecting units on a 2-D lattice with probability determined as a Gaussian function of physical proximity, and using a randomised small-world network connection strategy (Watts and Strogatz, 1998).

Small-world networks in this model are constructed in the form prescribed by Watts and Strogatz (1998) by firstly connecting each neuron to its closest K neighbours. Then, according to a probability of re-wiring $p(rewire)$, the connections between each unit and its two immediate neighbours are randomly assigned to other units in the network. Once each unit in the network has been considered, the neighbours two places away from each unit are then considered, and then those three places away, until each connection in the network has finally been randomly re-wired or left in place.

Clustering coefficients are obtained using the Watts and Strogatz (1998) measure, in which the clustering coefficient of a single neuron C_i is defined in terms of its neighbourhood of connected neurons V_i as the ratio of the number of connections (edges) e_{jk} that

exist between its neighbours $v_j, v_k \in V_i$, to the number of potential connections which *could* exist between its neighbours $|V_i|(|V_i| - 1)$:

$$C_i = \frac{|e_{jk} : v_j, v_k \in V_i|}{|V_i|(|V_i| - 1)} \tag{3.12}$$

The clustering coefficient of the whole network $\bar{C}$ is simply the average C_i over all neurons in the network:

$$\bar{C} = \frac{1}{N} \sum_{i=1}^{N} C_i \tag{3.13}$$

It has been shown that a process of diluting the synaptic weight matrix of a Hopfield network such that it is no longer fully-connected still causes it to behave in much the same manner as a fully-connected network when stored state vectors are generally of low activity (Evans, 1989). Examples of connectivity matrices for various connection strategies are shown later in figure 3.9.

3.2.6 Tau lesioning

Medical background

The tau hypothesis refers to the neurodegenerative effects of a modified (or *hyperphosphorylated*) form of the tau protein which aggregates with other fibres of tau and eventually forms the neurofibrillary tangles (NFTs) inside neurons which are prevalent in brains with AD. It proposes that the cognitive decline in AD is due primarily to loss of synapses and neurons (via a toxic form of the modified tau), and the subsequent loss of connectivity experienced (Spires-Jones et al., 2009).

The normal function of tau, of supporting the growth of neural cells via the construction of microtubules, and the transport of neurotransmitter-containing vesicles along the axon in mature cells, is disrupted in AD (Feinstein and Wilson, 2005; Johnson and Stoothoff, 2004). Instead of binding to the microtubules, tau in AD becomes sequestered into NFTs within the neurons (Ballatore et al., 2007) and as the level of normal tau in the

brain is reduced the microtubules disintegrate, causing further neuronal dysfunction. The existence of NFTs could also present a toxic gain-of-function by physically obstructing the transport of vesicles within the neuron (leading to cognitive impairment) and also by further sequestration of normal tau into the modified form as part of a cascade of neurodegeneration (Ballatore et al., 2007).

It is worth considering that, although the amyloid hypothesis provides a possibly more widely-encompassing view of AD, it has a number of significant unexplained problems, not least that "the number of amyloid deposits in the brain does not correlate well with the degree of cognitive impairment" (Hardy and Selkoe, 2002), and so the tau hypothesis and its relationship to the amyloid hypothesis remain an important subject for further research.

Computational implementation

Hopfield-type networks (including the T-F network) suffer from the inherent problem that the output layer is essentially the only layer of the network. Whenever lesions are applied to the network by deleting neurons, this necessarily results in the inability of the network to completely recover a cued pattern, regardless of the effects that the removal of these neurons may have had on the dynamics of the network, as the 'output layer' is now only partially complete and can no longer map with full accuracy to every given pattern. This results in a perceived decrease in network performance even though the network dynamics may not necessarily have changed.

Therefore, rather than performing total removal of neurons or deletion of synapses, it is suggested that results more representative of AD pathology could be obtained by the introduction of a subtle shift from *deletion* to *atrophy*, whereby whole groups of synapses with a single neuron at their centre are affected in a similar way at the same time. In order to model more specifically the pathological effects of tau NFTs, the output of selected neurons could be partially muted to simulate the effects of axonal blocking and microtubule collapse by NFTs.

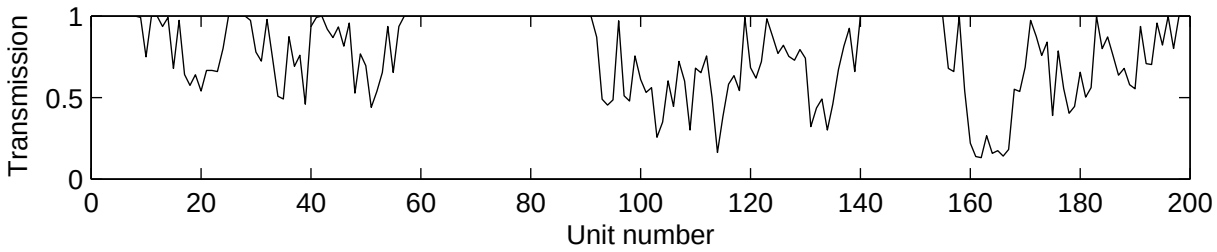


Figure 3.3: Characteristic plot of activity transmission levels after spatial tau lesioning.

To simulate the sequestering of hyperphosphorylated tau and the subsequent cascading spread of damage, neighbouring neurons could also be muted by a slightly smaller amount, with new tau lesion centres subsequently formed near to existing lesions, and the resulting distributed damage occurring in less of a severe 'binary' manner as with random deletion of synapses or neurons.

In computational terms, lesioning can be performed in steps of size d (equivalent to the deletion step size used during synaptic or neuronal deletion). The locations of the centres of the first set of lesions (the tau *seeds*) are chosen at flat random from across all neurons and are assigned to set M. Subsequent lesioning steps proceed as follows:

- A subset $m \subseteq M$ of size d locations are chosen at flat random. With Gaussian probability centred on each element in m, a new set m' of neighbours is chosen.

- Each neuron in the network has its transmission of incoming activation dampened by multiplying postsynaptic potentials by a value drawn from a Gaussian distribution as a function of the neuron's proximity to the lesion centre (i.e. $x - \mu$), and with lesion width σ (suggested as 2), such that those neurons closest to the lesion centres in m' are most heavily diluted, and those distant from the elements of m' are relatively unaffected.

- The new lesion centres m' are added to M, and the process is repeated with the now larger set of lesion centres M. This results in characteristic lesioning as seen in figure 3.3.

3.3 Results

Experiments consisted of repeated iterations of a process involving deletion of a fixed number Δd of neurons in the network, calculation of new synaptic compensation values, and finally testing the network's recall performance over the stored set of patterns. Unless otherwise stated, all experiments were performed in a network with the following parameter values. All values are the same as those used in Ruppin and Reggia (1995a), with the exception of the increased network size and the partial connectivity:

Network size $N = 1600$, connections per unit $K = 200$, neural threshold $\theta = 0.048$, noise $T = 0.005$, learning rate $\gamma = 0.025$, external input strength (learning mode) $e_l = 0.065$, external input strength (retrieval mode) $e_r = 0.035$, coding rate $p = 0.1$, deletion step $\Delta d = 0.01$. Results were averaged over 10 runs and the number of patterns stored on each run was 10, with graphs showing mean and standard deviation across the 10 runs.

3.3.1 Random deletion and local field-dependent compensation

In the first experiment, an attempt was made to replicate the results of Ruppin and Reggia (1995a) using the improved neural field-dependent compensation rule of Horn et al. (1996), in which compensatory mechanisms extended the working life of the network during repeated synaptic deletion. The network was connected with a Gaussian connection strategy and a set of 10 patterns was stored in the network. The average retrieval success rate (*overlap*) after various levels of deletion was plotted. The results shown in figure 3.4 are qualitatively the same as those shown by Horn et al. (1996), and can be used as a performance baseline for later experiments.

3.3.2 Compensation using recent versus remote memories

Ruppin and Reggia (1995a) observed a gradient of damage by repeating a process of learning a set of patterns then subsequently deleting a proportion of the connections between units, over several rounds of learning and deletion. Their results, based on a

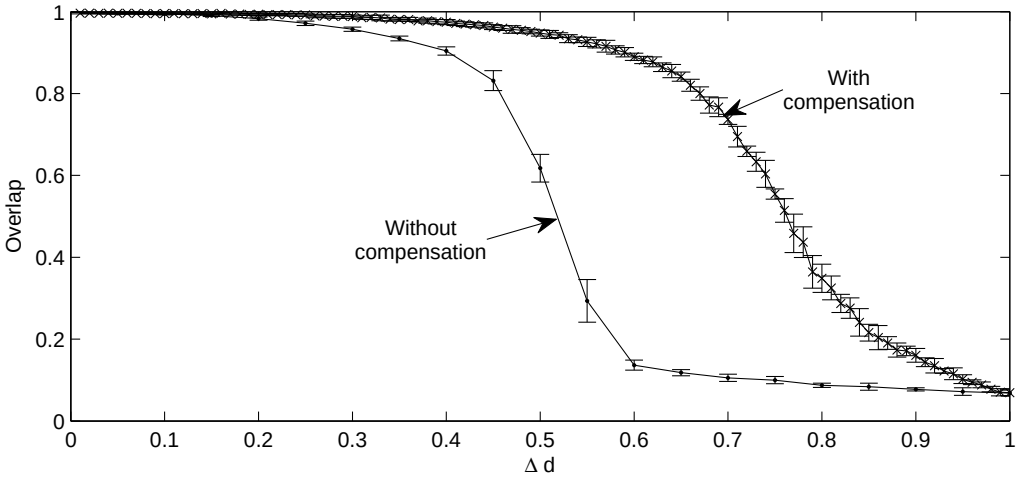


Figure 3.4: Performance over synaptic deletion without compensation (leftmost curve) and with (rightmost curve). The deletion step size Δd is larger for the curve without compensation, but this does not affect the overall result as deletion step size-sensitivity is only introduced with compensatory synaptic scaling (i.e. the results are linear with respect to d.

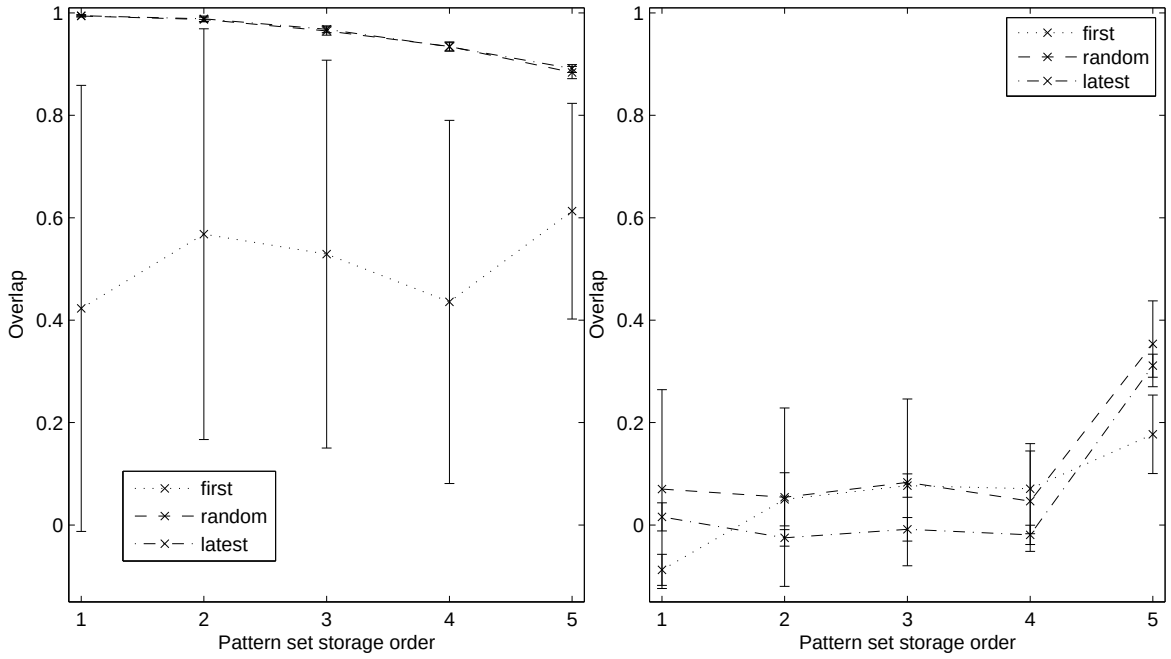
fixed compensatory scaling strategy, showed a clear decrease in recall performance for patterns learned recently compared with those learned earlier in the process.

This is due to the compensatory mechanism widening the basins of attraction of stored patterns in synaptic weight space; those patterns which were stored earliest naturally receive the most compensation, and therefore obtain the widest basins of attraction, causing the network to fall more easily into these states[1]. The results of Ruppin and Reggia (1995a) were replicated, and are shown in figure 3.5.

As the local field-dependent compensation strategy of Horn et al. (1996) works by using the retrieval of stored memories for comparing average post-synaptic potentials before and after damage, the choice of memories which should be used for this purpose becomes significant due to the different retrieval success rates of patterns stored early in the lesioning process compared to those stored more recently.

As shown in figure 3.5a, if only remotely-stored patterns are used during the compen-

[1]Models of schizophrenia based on the same T-F network have shown the logical end result of this process: that is, a much-reduced set of retrievable patterns with very large basins of attraction into which the network spontaneously and repeatedly falls during learning and retrieval. The authors suggest this as a mechanism for schizophrenic hallucinations and fixations upon a particular and limited set of ideas (Horn and Ruppin, 1995). It is also interesting to note that episodes of psychotic hallucinations are increased in Alzheimer's disease (Savioz et al., 2009).



(a) Final deletion = 0.35 (b) Final deletion = 0.45

Figure 3.5: Performance on separately stored sets of memories. The network was alternately presented with sets of 6 patterns then subjected to a process of deletion with compensation using only the first set of patterns stored (dotted line), only the last set of patterns stored (dot-dash line), or using a random set of 6 patterns drawn from all those previously stored (dashed line). By the final round, the total proportion of deletion was either 0.35 (figure 3.5a) or 0.45 (figure 3.5b).

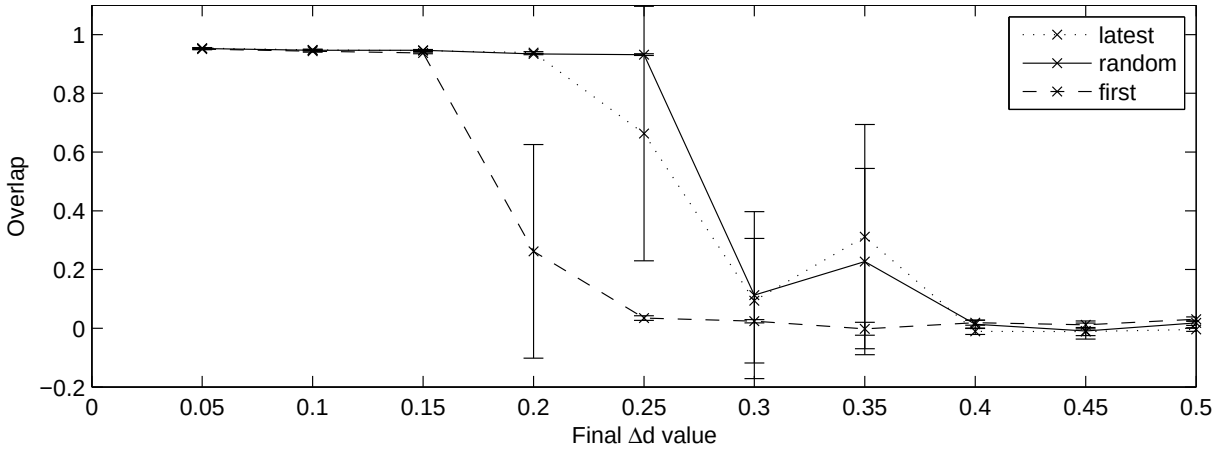


Figure 3.6: Average performance on all learned sets (5 sets of 10 patterns, $N = 1200$, $K = 200$) after different final levels of deletion. First-set compensation (dashed line) performance declines significantly earlier than latest-set compensation (dotted line), which itself is marginally less robust than random-set compensation (solid line).

satory process (dotted line) the performance of the network is severely degraded even at a relatively low level of deletion, whilst use of the most recently-stored patterns during compensation is almost indistinguishable in performance from the results when using a random set of patterns drawn from all those previously stored (dot-dash and dashed lines, respectively).

In these cases, a clear gradient of learning has been observed such that patterns stored most remotely are recalled more successfully than those stored more recently, due to their larger basins of attraction within the weight space. (Note that this is a distinct phenomenon from serial-position effects in which recency and primacy of items within a list correlate with greater recall, as the network is storing time-separated sets of patterns in between periods of damage rather than a single list of items). At higher, catastrophic levels of deletion (figure 3.5b), compensation using any sets of patterns results in a similar low level of recall performance. Clearly, despite the fact that remotely-stored patterns are the most successfully-retrieved, use of only the remotely-stored patterns during compensation results in much earlier decline of the network performance as deletion progresses (figure 3.6). Conversely, compensation using the most recently-stored patterns, or sets of patterns drawn at random from all the available patterns, results in greater robustness to damage.

When reducing the scope of the sets of patterns used in the compensation process in this way, the compensation process becomes more susceptible to noise, as the number of unique trials which may be performed to estimate neural connectivity levels is lower than it would be if all of the patterns were able to be used in the trials. This is exactly the same effect as reducing the number of repetitions of an experiment, and getting larger error bars as a result. This has implications for sufferers of Alzheimer's disease: if the network is damaged to such an extent that the gradient observed in figure 3.5a (the uppermost dashed and dot-dash lines) is evident, but the network is not yet catastrophically damaged, then there may be a greater likelihood that compensatory mechanisms will activate remotely-stored patterns compared to recently-stored ones, due to their larger basins of attraction and thus their higher recall success. As this experiment has shown, this could actually

lead to earlier overall decline of cognitive abilities, and a cycle of correspondingly worse recent memory retrieval performance and thus continued use of a dwindling set of remote memories during compensation. Additionally, this finding places the deletion threshold of the model (beyond which all recall is severely affected) closer to the reported 10-30% atrophy levels seen before the symptomatic damage evident in AD.

3.3.3 Connection strategies

Effects on network capacity

Next, the effects on network capacity and robustness to damage of various connection strategies were compared. Firstly, networks were created with $N = 800$ units with connection density $K = 0.125N$. This was a smaller network size than in the previous experiments, in order to allow a greater number of runs (50, in this case) to be undertaken for each network type. The networks were wired with Gaussian, flat-random, and small-world (with various values for $p(rewire)$) connectivity. Patterns were stored in each network according to equation 3.1 and retrieved immediately after storage. The average retrieval success rate was plotted against the small-world clustering coefficient of the network's connection matrix, and when the average overlap measure dropped continuously below 0.8, the network was assumed to have reached its capacity (figure 3.7).

The results indicate that specific network connectivity generally has little effect on capacity, with the random and Gaussian networks appearing on the same trend as the small-world networks, but it appears that network capacity is significantly reduced in more highly-ordered networks with high clustering coefficients such as small-world networks with low values for $p(rewire)$. Although each network structure contains the same number of connections, and should therefore have effectively the same capacity, the difference becomes clear when examining the layout of the connections and the related small-world clustering coefficients. Compare figures 3.9a and 3.9b, both of which were constructed using the small-world algorithm. The majority of connections in the network

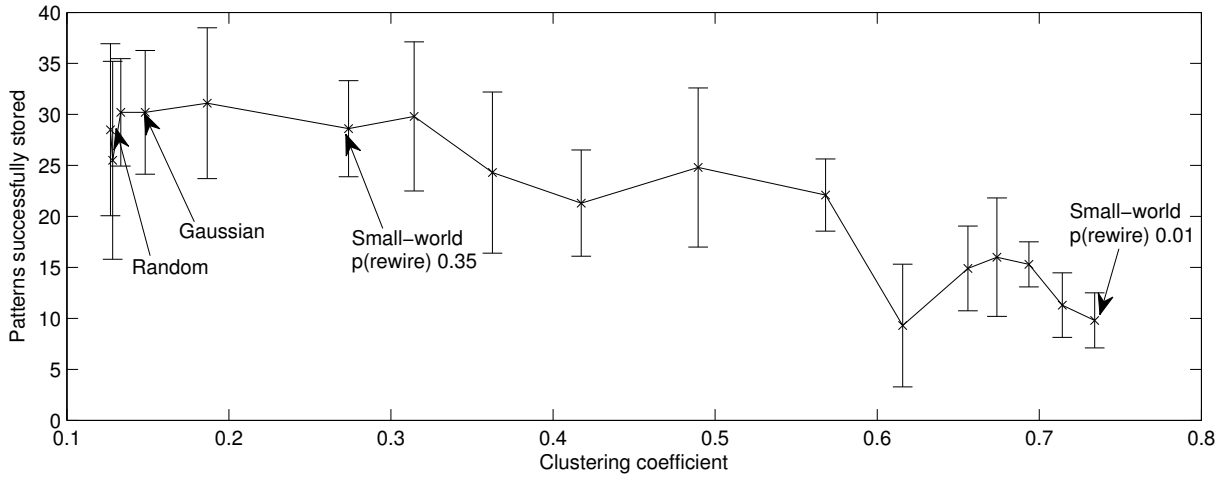


Figure 3.7: Capacity in networks with different small-world clustering coefficients. Flat-random, Gaussian and representative small-world networks are indicated individually.

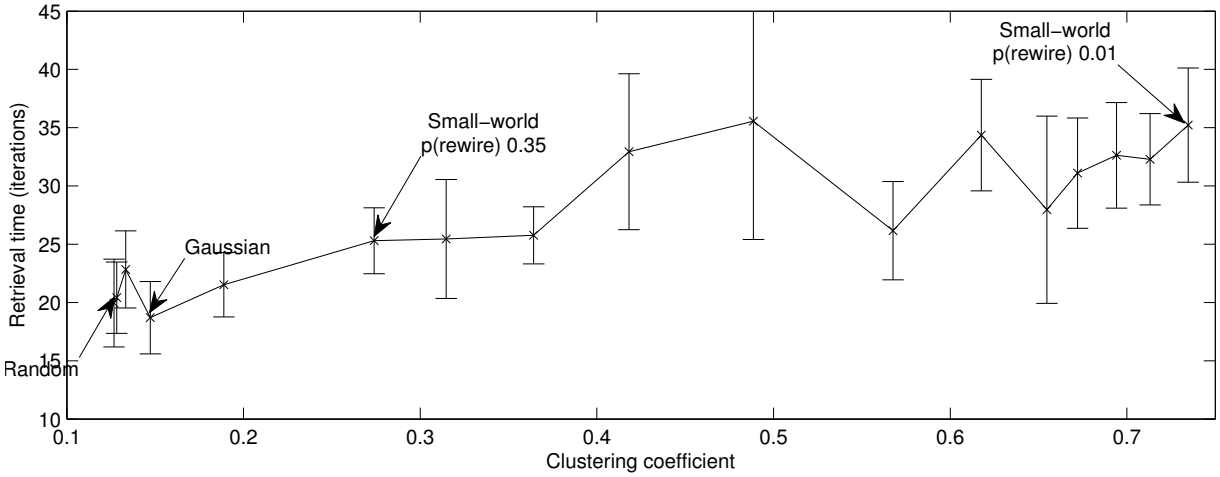


Figure 3.8: Retrieval times in networks with different small-world clustering coefficients.

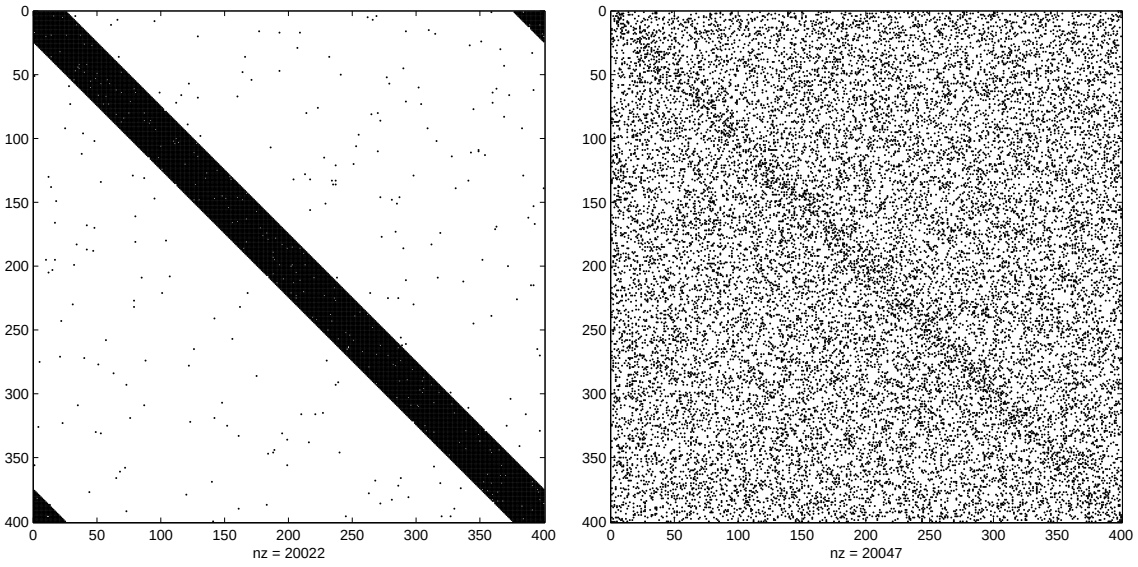
with $p(rewire) = 0.01$ are located incredibly densely within the local neighbourhood of each neuron, with only a few projections to more distant parts of the network, resulting in a clustering coefficient of 0.73.

This appears to have two effects: the first becomes clear when considering the pattern recall times in figure 3.8, which shows that in highly-regular networks the number of iterations required for the network to fall into a stable state is higher. This is likely to be due to the lack of distant projections to other parts of the network: activation is 'slowed-down' by having to flow through a closely-linked chain of units from one extent of the network to the other, whilst a network with less regularity and more distant projections (as seen in figure 3.9b) can effectively take short-cuts when activation to distant parts of the network is required. If this activation degrades over time as it traverses the network in small steps, or if there is a limit to the permitted time between cueing and retrieval of a pattern, it is clear to see that these effects could result in greater retrieval failure rates (and thus lower effective capacity) than in a network with less regularity in its connection matrix.

Effects on redundancy and robustness

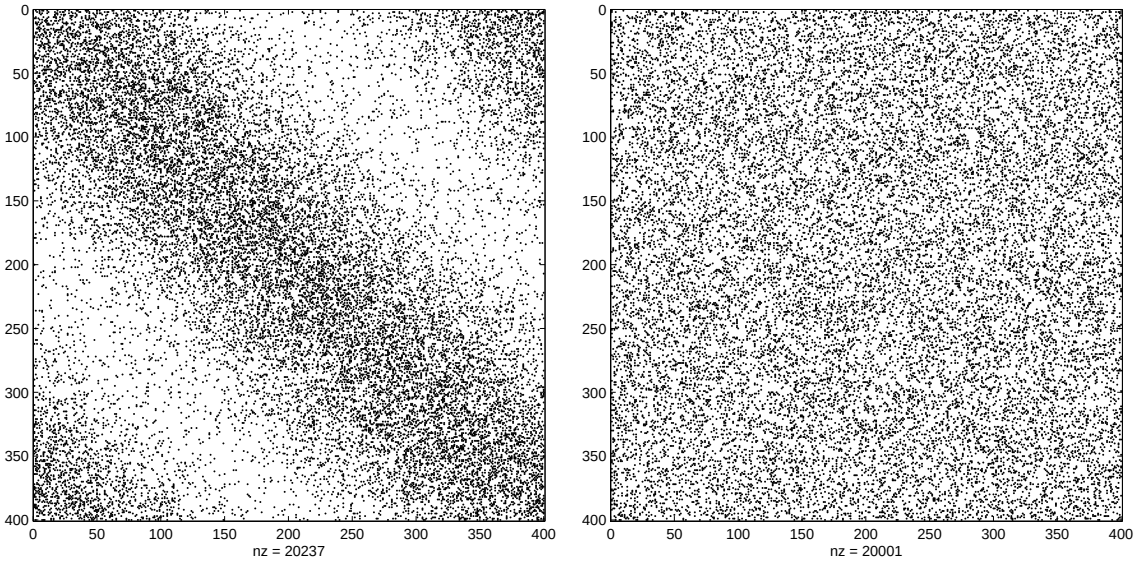
The second effect concerns the information capacity of the connections in the network. The high density of local connections in the regular network leads to synaptic redundancy, as activation between any two nearby neurons can take multiple paths between them. Necessarily, redundancy where more than one connection carries the same information results in a reduction in information capacity elsewhere in the network, as previously shown, but increased redundancy should also lead to networks which are more robust to damage.

To test this prediction, a profile of deletion (without compensation) was obtained for networks connected with small-world ($p(rewire) = 0.01$) and flat-random connectivities in 1600-unit networks with connectivity $K = 0.125N$. The resulting plot in figure 3.10 shows a marginally smoother rate of decline and greater longevity of performance in the small-



(a) Small-world (0.01)

(b) Small-world (0.9)



(c) Gaussian

(d) Flat-random

Figure 3.9: Connection matrices of networks connected with small-world, Gaussian, and flat random strategies ($N = 1600$, $K = 200$, $p(rewire) = 0.01, 0.9$, only every fourth connection plotted for clarity).

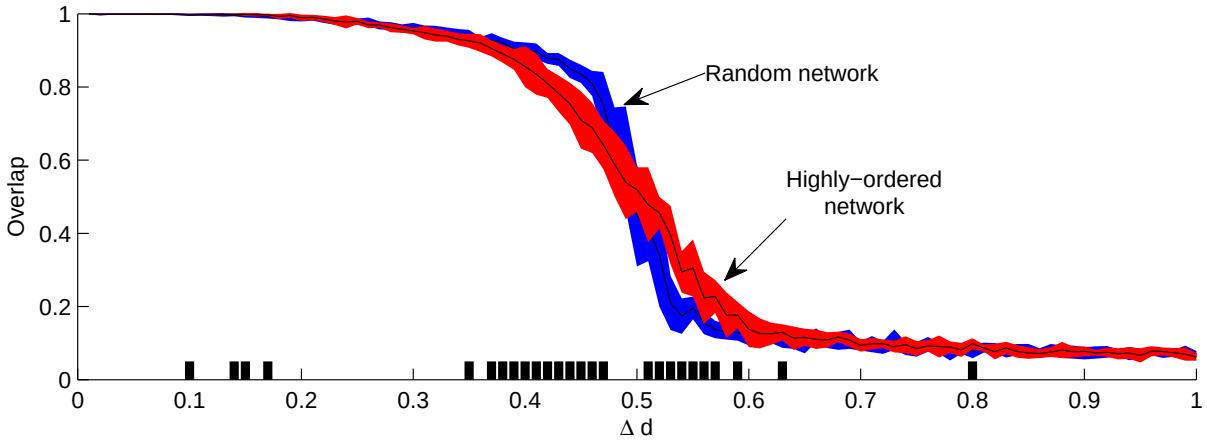
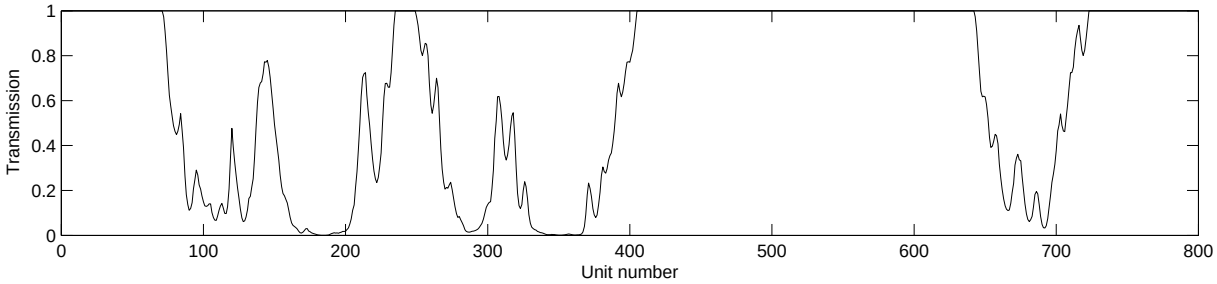


Figure 3.10: Deletion without compensation in random (blue) and highly-ordered (red) networks. Locations of statistically-significant differences highlighted by markers.

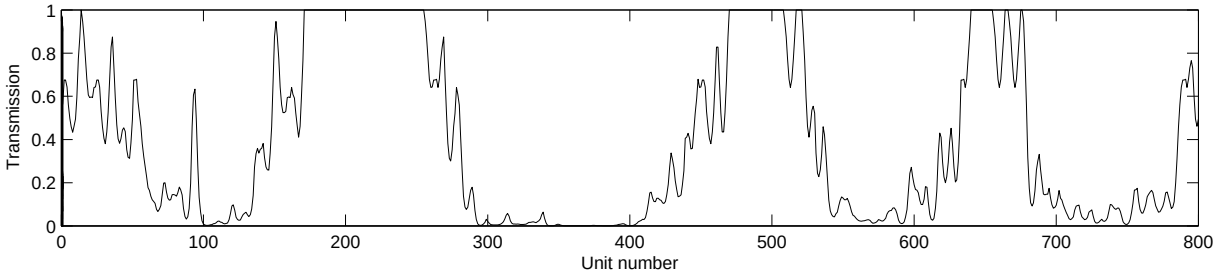
world network than in the random network, indicating that the high local connectivity density does indeed lead to redundancy and hence greater robustness to damage, but at the expense of lower capacity. The differences were measured statistically using Student's t-test, showing significant ($p < 0.05$) differences close to the middle section of deletion. Nevertheless, the effects are relatively small overall.

3.3.4 Tau lesioning

To identify the changes in behaviour when more distributed, variable-rate tau damage occurs within the network, a network with $N = 800$ units was connected in a Gaussian manner and tau lesioning was performed according to the method described in section 3.2.6, with random-set compensation. Two rates of tau lesioning were inspected: in addition to the standard rate in which the neuronal outputs were muted by an inverse Gaussian probability as a function of the neuron's distance from the lesion centre, a second rate was tested in which the muting amount was squared so as to increase the speed with which the lesions resulted in full neuronal blocking, and the width of the distribution used for choosing new nearby tau lesion centres was doubled. Examples of the resulting comparable increase in lesioning can be seen in figure 3.11. A further, currently untested, method of altering the tau lesioning rate would be to consider each unit more than once



(a) Standard-rate tau lesioning



(b) Enhanced-rate tau lesioning

Figure 3.11: Lesions applied to example networks after every unit has been considered once. In the second graph, the rate of reduction of neural transmission in each lesioning step has been squared and the horizontal spreading speed of the lesioning has been doubled.

until full blocking of all units occurs.

The results in figure 3.12 show a very different profile to basic deletion (see figure 3.4). Rather than a smooth decline in performance which tails off towards zero, a sudden catastrophic decline in performance occurs during a single step of tau lesioning. The performance then steadily recovers, but only to a fraction of the original performance, as the compensatory mechanisms attempt to "catch up" with the sudden decrease in activation. As seen in figure 3.11, there are still areas of the network which are undamaged (transmission remains at 1), and it is likely that it is these areas which contribute to the above-zero final performance of the network.

Although the precise timing of the sudden decline varies randomly between test runs, each line on the graph traces essentially the same shape. Indeed, it was found that the differing rates of neuronal damping shown in figure 3.11 resulted in exactly the same profile of altered performance during lesioning, except that the performance drop-off was experienced correspondingly earlier or later with faster or slower tau lesioning rates.

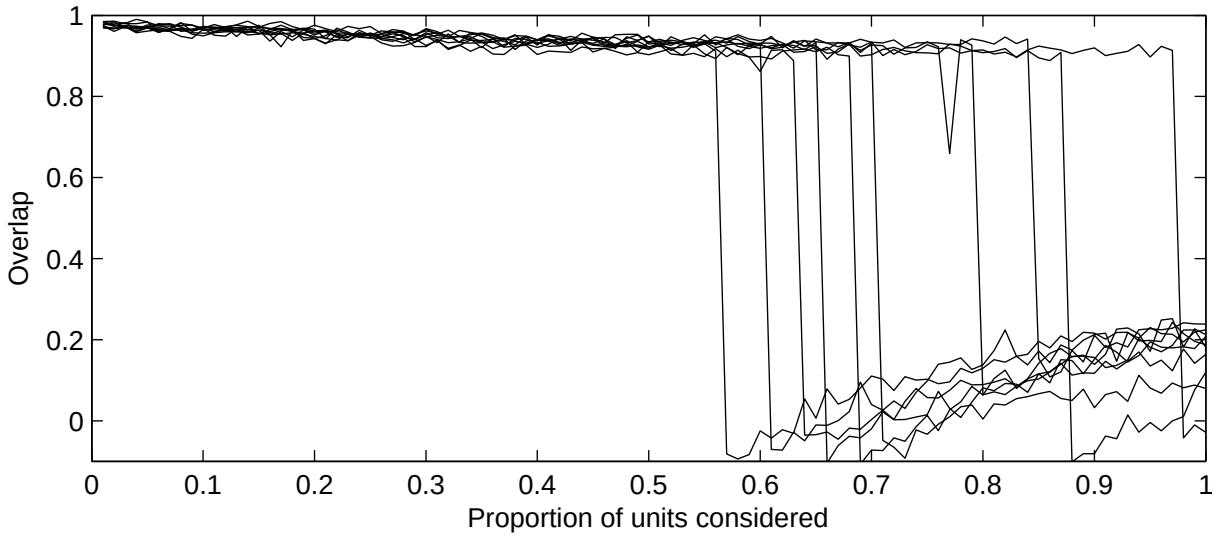


Figure 3.12: Performance plotted against number of units considered for lesioning with tau (enhanced rate). Due to the large differences in the time of onset of impairment, individual test runs have been plotted for easier comparison.

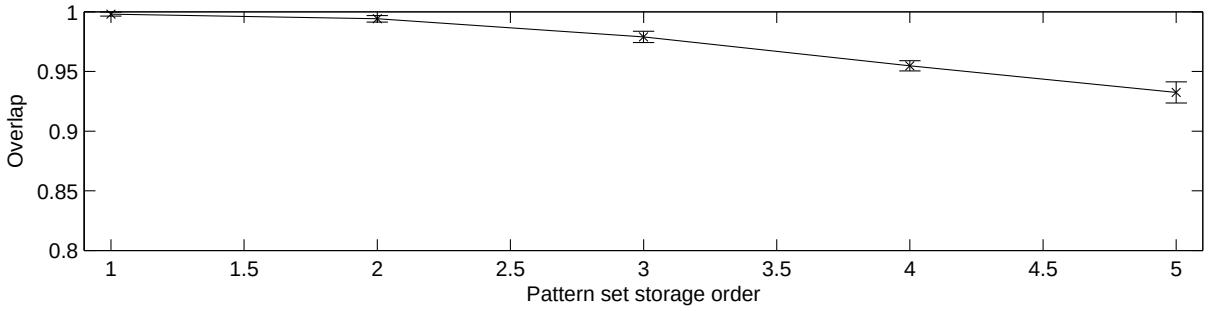


Figure 3.13: Gradient of performance on sets of patterns stored at different stages of lesioning with tau.

To test that the observed gradient of learning in sets of patterns over time (figure 3.5) still occurs with tau lesioning, the experiment in section 3.3.2 was re-run in a Gaussian-connected network with tau lesioning instead of deletion. The patterns used for compensation were drawn at random from all those previously stored. The results in figure 3.13 are comparable with those in figure 3.5, indicating that tau lesioning in this way does not destroy the effect of reduced retrieval of recent compared to remote patterns.

3.4 Discussion

This chapter has presented updates to a long-standing and widely-cited computational model of Alzheimer's Disease (Ruppin and Reggia, 1995a), including first successfully replicating, and then extending, the experiments of Horn et al. (1996) on the effects of local, field-dependent compensatory synaptic scaling mechanisms within the model.

The differing effects of using recent, remote, and random sets of memories to calculate compensatory signal terms has been shown, revealing that the network is sensitive to the choice of memories used. Using only remote memories to calculate the signal term results in reduced variation and therefore greater noise within the compensatory mechanism, and an earlier decline in performance as synapses are deleted (closer to the $10-30\%$ range seen in AD patients (Minati et al., 2009), although the exact level of performance depends also on number of patterns stored and network capacity). The implications for AD patients are shown in the context that initial retrieval of remote memories at early stages of damage is actually more reliable than with recent memories: if the brain makes use of this effect and uses the more readily-available remote memories to calculate compensation, not only do the recently-stored memories continue to become less reliable than the remote memories, but the noise in the system leads to earlier onset of catastrophic decline.

Of course, the idea of presenting a complete, noise-free pattern during compensation, obtained from the repertoire of stored patterns, is biologically implausible, as there is no fixed external repertoire of patterns. In the brain, the patterns are instead likely

to be spontaneously retrieved by noisy activation of external inputs to the network, for example during dreaming (Horn et al., 1996). This means that although patterns may be activated at random during the compensation process, the larger basins of attraction of the remotely-stored patterns will cause the network to fall into them more often, thus reinforcing these patterns during the compensation process, and further reducing the size of the set of patterns available for recall, leading to noisier compensation and progression of damage.

These results suggest the possibility that damage progression could be slowed therapeutically by encouraging patients to focus on recalling memories and states acquired throughout that individual's lifetime, including recently, rather than focussing on the easier-to-recall memories from (relatively) early life, thus selectively enlarging the basins of attraction of recent memories and making a greater variation of patterns available to the compensatory mechanisms.

It has also been shown that network capacity and resilience is related to the regularity of connections within the network. High small-world clustering coefficients lead to redundancy within the network, meaning greater resilience to damage but at the expense of lower capacity, as well as longer pattern retrieval times. This is consistent with the findings of Supekar et al. (2008) who examined small-world functional networks in the brain and found a key correlation between loss of small-world connectivity and onset of AD symptoms. It is also worth considering that the presence of redundant networks in the brain capable of withstanding damage is consistent with evolutionary constraints. Further examination of the relationships between small-world clustering, robustness, retrieval speed and network capacity could be revealing, as well as studies into how this operates within the principle of neural Darwinism (pruning of weaker synapses during brain development).

Lesioning with simulated tau rather than standard synaptic deletion has been shown to create a very different profile of damage by allowing all neurons and synaptic connections to remain present (so output patterns are not artificially altered) and instead damping

inter-neuronal transmission. Whilst initially offering a much more graceful decline in performance due to the persistence of synaptic connections and output units, consistent with the slow degradation seen in AD, the drop-off in performance when it finally occurs is much more severe with tau lesioning than with synaptic deletion despite some later compensatory recovery of performance.

Further work will be needed to ascertain whether this deletion profile offers a more plausible explanation of AD symptoms and whether the observed temporary improvement in recall after some level of catastrophic damage can be medically corroborated, but it must be borne in mind that tau pathology represents only a subset of the processes underlying AD. It would be beneficial to extend this concept and show in a similar way the effects of alternative AD pathologies, such as the beta-amyloid hypothesis.

The Tsodyks and Feigel'Man (1988) model studied in this work is only a basic associative network with limitations in processing ability and a relatively constrained range of behaviour. More sophisticated artificial neural network-based models such as LEABRA (O'Reilly, 2001), networks of spiking neurons (Gerstner and Kistler, 2002), and reservoir networks (Lukosevicius and Jaeger, 2009) are available and could provide further insights into the effects highlighted in this chapter. In particular, it would be interesting to examine the effects of compensatory scaling and connectivity strategies within a reservoir computing framework due to the large potential for exploration of the currently poorly-understood dynamics, and the greater computational power (potentially offering the representation of more varied symptoms of AD than simple pattern recall) of these systems.

3.5 Chapter summary

In this chapter, a basic Hopfield network model, developed by Tsodyks and Feigel'Man (1988) and enhanced by Ruppin and Reggia (1995a), was examined, and the effects of Alzheimer's disease-related damage on networks with various different connectivity and

compensatory strategies were tested.

The effects of using recent versus remote memories during compensation were shown, indicating that compensatory mechanisms may induce a feedback loop of reinforcement for easily-retrieved remote patterns ("old memories"), at the expense of weakening recently-stored patterns. This reduction in variation of possible input to the compensatory mechanism leads to noisier compensation, and earlier failure of the network. Therapeutic strategies are suggested which may reduce the progression of damage. This is the first time such effects have been shown in a computational model and explained.

The effects of connectivity density and wiring strategy on network capacity and robustness to damage were characterised, with strong correlations between loss of clustering coefficient in small-world networks and degradation of network performance. Finally, an alternative lesioning process representing tau pathology was shown to lead to a very different damage profile to the observed effects of traditional random or spatial synaptic and neuronal deletion.

CHAPTER 4

INFORMATION-SELECTIVITY OF SCALING-DRIVEN PATHLOGY

Chapter synopsis

This chapter looks deeper into the workings of the model at the individual neuronal level under conditions of damage and compensatory synaptic scaling, and develops the hypothesis of *information-selectivity*. This is the idea that, via a side-effect of compensatory synaptic scaling, Alzheimer's disease may selectively and aggressively attack the least important neurons in the network first, thus delaying the onset of cognitive symptoms whilst the disease spreads, thus disguising the presence of the disease and making timely intervention and treatment significantly more difficult.

Firstly, it gives the background to Small (2008)'s hypothesis of the progression of Alzheimer's disease driven by compensatory synaptic scaling. We show how this may be modelled in an artificial neural network simulation, using the Tsodyks and Feigel'Man (1988) model from the previous chapter as an example. Next, it briefly introduces the well-established field of *information theory* and how it can be applied to neural systems to obtain a measure of "significance" for each neuron. These tools are used to outline the theory of information-selectivity and how it relates to Small's progression hypothesis. Experiments are then performed to demonstrate the effects of information-selectivity on network performance, for pattern retrieval accuracy and timeliness. Finally, we discuss

the results in the context of utilisation of available associative storage capacity within the brain.

Work derived from this chapter has previously been published as Rowan (2012).

4.1 Introduction

Prevailing medical opinion holds that the density of plaque deposits of beta-amyloid (Aβ) in the brain does not necessarily correlate well with cognitive decline in Alzheimer's Disease (AD), particularly in the early stages of the disease (Hardy and Selkoe, 2002; Savioz et al., 2009; Minati et al., 2009), yet the *amyloid hypothesis* remains the key investigatory path in AD research due to its undeniable significance to the overall pathology of the disease (Hardy, 2009). At the same time, it is unclear what mechanisms drive progression of beta-amyloid throughout the network. Of particular interest on this topic is a hypothesis proposed by Small (2008) which describes the progression of AD as a self-reinforcing cascade resulting from synaptic scaling within the network as a response to damage.

Synaptic scaling is a mechanism by which neurons can compensate for reductions in their postsynaptic potential as connected contributory neurons die off or synapses between the neurons are blocked (e.g. by beta-amyloid, cell death, or tau neurofibrillary tangles (Ballatore et al., 2007)). In an un-compensated network, average postsynaptic potentials into each neuron would decrease with damage and firing thresholds would not be reached, leading to overall failure of the network. Synaptic scaling mitigates this effect by increasing the weights of the remaining connections to each neuron, resulting in a maintained average postsynaptic potential profile albeit with increased noise (Savioz et al., 2009; Small, 2008; Horn et al., 1996). Small (2008) proposes that this synaptic scaling mechanism directly drives progression of AD within the network. It should be noted that alternative computational mechanisms for mediating presynaptic activity exist (such as Oja's rule (Oja, 1982) and the modifiable firing thresholds of BCM neurons (Bienenstock et al., 1982)). However, regardless of the precise mechanism of activity regulation, the

same principle of excitotoxicity due to increased excitability applies.

In the previous chapter, we introduced updates to an early Hopfield-class attractor neural network model of some of the symptoms of AD (Ruppin and Reggia, 1995a) which incorporated local, field-dependent synaptic compensation (Horn et al., 1996). This chapter further updates the model to include simulated Aβ pathology according to Small's hypothesis of disease progression (Small, 2008) and examines the effects, using an information-theoretic approach, of Aβ on the cognitive performance of the network.

4.2 Background

4.2.1 Beta-amyloid

The amyloid hypothesis states that plaques consisting of extracellular deposits of beta-amyloid (Aβ) derived from amyloid precursor protein (APP) are formed in the brain, leading to neuronal toxicity, cell death, and subsequent neurodegeneration and cognitive deficiencies (Hardy and Selkoe, 2002).

It is believed that amyloid pathology affects several aspects of the overall process of AD, altering neurological processes directly by blocking synapses (Hardy, 2009) as well as exacerbating the effects of tau and cholinergic pathology (Wilcock et al., 2009).

However, original predictions that Aβ should be neurotoxic have been found lacking in support (Hardy, 2009). One reason for this is the observed discrepancy between the number of solidified amyloid plaques found in the brain, and the level of cognitive decline observed in the early stages of Alzheimer's disease (Hardy and Selkoe, 2002; Savioz et al., 2009; Minati et al., 2009). A possible mechanism for neurotoxicity, via calcium signalling systems, is believed to arise from the interaction of Aβ with neuronal membranes, leading to relaxed regulation of Ca^{2+} flux into the neuron and the subsequent increase in Ca^{2+} ions inside the neuron beyond a critical level at which apoptosis (programmed cell death) (Hynd et al., 2004), or even necrosis (traumatic cell death) occurs (Berridge et al., 2000).

The modification of Ca^{2+} signalling systems is also believed to underlie some of the cognitive deficits seen in AD more directly (Berridge, 2010). Ca^{2+} is closely related to the action of glutamate receptors on neurons in the hippocampus which are essential for long-term potentiation as part of a Hebbian learning process (Clapham, 2007), implying that disruptions in calcium signalling systems may lead to learning and memory deficiencies seen in AD. "Altered Ca^{2+} signaling accelerates Aβ formation, whereas Aβ peptides, particularly in soluble oligomeric forms, induce Ca^{2+} disruptions. A degenerative feed-forward cycle of toxic Aβ generation and Ca^{2+} perturbations results" (Demuro et al., 2010).

As well as forming external amyloid plaques, there is evidence that Aβ accumulates internally within neurons, inhibiting proteasomes (protein complexes which regulate the concentration of proteins within a cell and clear up pathological proteins), thus leading to a build-up of tau protein, and subsequent tau pathology. Proteasome inhibition also leads in turn to a further build-up of Aβ, thus exacerbating the cascade, and to calcium dysfunction. It has been suggested that intracellular Aβ oligomers can adversely affect long-term potentiation in the hippocampus, leading to further cognitive decline (LaFerla et al., 2007).

4.2.2 Synaptic scaling and disease progression

Small (2008) proposed a mechanism by which progression of Aβ pathology in Alzheimer's disease may be tightly coupled with synaptic compensation, in a self-reinforcing cycle of neurodegeneration. This hypothesis has a key feature in common with a much earlier work by Wallenstein and Hasselmo (1998), chiefly that the brain's own neuroregulation mechanisms have some role to play in selective spreading of the disease pathology throughout the brain. Whilst Wallenstein and Hasselmo concentrate on the spread of "runaway synaptic modification" (pathological exponential growth of large numbers of synaptic connections) progressing along paths of high activity, Small proposes a similar process involving the spread of Aβ pathology.

In brief, Small proposes that initial Aβ-related neurotoxicity in a single neuron causes a responding increase in synaptic compensation amongst connected neurons, necessary for each to maintain its average postsynaptic potential profile after the death of the Aβ-affected neuron. This increased synaptic compensation effectively increases the excitability of the neurons concerned, leading to locally raised intracellular calcium levels in the sub-network to which the toxic neuron was connected. As calcium plays a key role in mediating Aβ neurotoxicity (Demuro et al., 2010), these neurons now gain increased vulnerability to the surrounding pathology: a self-reinforcing, locally-spreading cycle now occurs, whereby synaptic scaling in healthy neurons which were locally connected to an Aβ-toxified neuron leads to the spread of the pathology throughout the network.

4.3 Information-selectivity

4.3.1 Considering connectivity

The scaling-driven progression proposed by Small (2008) raises an interesting question regarding the significance of which neurons succumb to the pathology, and at which point during the lesioning process they do so.

From within the context of synaptic compensation, one can easily deduce that neurons which are only sparsely connected to others will suffer a large relative drop in average postsynaptic (input) potential when one of their connected input neurons dies, and therefore a large compensation factor must be applied to make up for this shortfall. Conversely, a neuron which is densely and strongly connected to its neighbours will suffer much less severe a decrease in postsynaptic potential for each lost incoming connection, and its subsequent compensation will be low, safeguarding the neuron from excitability-induced pathology (figure 4.1).

Due to their greater compensatory rates, sparsely-connected neurons are thus more likely to succumb to the pathology early in the progression of the disease.

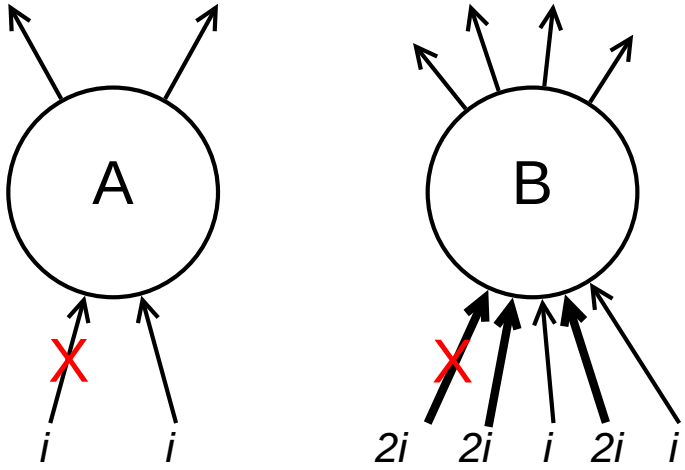


Figure 4.1: Demonstration of the effects of connectivity density and strength upon compensation rates. Neuron A has weak, sparse input activation totalling some arbitrary value $2i$. If one of these input connections is deleted (e.g. after a connected neuron dies; denoted with a red cross), the remaining connectivity (see equation 3.7) $w' = \frac{1i}{2i} = 0.5w$, and the resulting compensation factor $c' = \frac{1}{w'} = 2c$. Neuron B has strong, dense input activation totalling $8i$. Deletion of one of these input connections (here, one of the stronger inputs, but the effect is even more pronounced when one of the weaker inputs is deleted) gives remaining connectivity $w' = \frac{6i}{8i} = 0.75w$ and lower resulting compensation $c' = \frac{1}{w'} = 1.333c$.

This principle is independent of the specific compensatory mechanism employed within the model, providing that neurons have some mechanism of estimating their remaining connectivity $\hat{w}_i$ and scaling this connectivity up by some factor c_i in order to maintain an average postsynaptic potential.

According to the Hebbian principle, the development of strong, dense input connectivity to a given neuron during the learning process indicates some kind of significance of that neuron to its local network ("*cells that fire together, wire together*", or more formally, connection strengths are modified according to the degree of correlated activity between input and output). Additionally, it is known that during learning, "network connections develop in such a way as to maximize the amount of information that is preserved when signals are transformed" (Linsker, 1988), and thus we can pose the questions:

- Is there a possible selectivity of AD pathology for neurons with low information contribution into the rest of the network?

- If so, could the preference of pathology for low information-contributing neurons explain the observed poor correlation between Aβ deposits and cognitive decline in the early stages of AD?

4.3.2 Significance of a single neuron

If it can be assumed that the development of strong, dense input connectivity to the neuron via a Hebbian learning process indicates some kind of significance within the neuron's local network, then it would seem intuitive that other neurons might weight the output of this neuron significantly as well.

To find the significance of an individual neuron to the rest of the network, we must consider both the *number* of its outward connections to other neurons as well as the *strength* of these connections, and also the *variability in its firing state* over time. Linsker explains in detail why it is that a neural system should seek to maximise each neuron's variance, and that variance implies information (Linsker, 1988). One intuitive illustration of this principle is that if a neuron only ever exists in one state (e.g. it is firing or quiescent in all cases), then its output variance is zero over all inputs, and it contributes no useful information to the network. This principle is rooted in information theory and explained in further detail by Borst and Theunissen (1999).

The *significance measure* of each neuron i under random network update conditions, in which the network is initialised with a random state and allowed to update, can be calculated as follows, where N is the number of neurons in the network and $S_i \in \{0, 1\}$ is the firing state of neuron i:

- For m repetitions per neuron i (suggested $m \geq N$):
- Set the network to a random state satisfying the coding rate $p \ll 1$ (here, $p = 0.1$).
- Allow each neuron to update according to the update rule in equation 3.11.
- Obtain the outgoing presynaptic potential to each of its afferents j (neurons which

i activates):

$$h_j^m = \sum_j W_{ji} S_i \tag{4.1}$$

This means that h_j varies at each update with the values of S_i (the neuron's firing state) at each update.

- The signifcance Υ of neuron i to the network is taken as the variance of h_j over all m repetitions:

$$\Upsilon = Var(h_j^m) \tag{4.2}$$

This has the result that a neuron with a large number of outgoing connections and/or high firing variance is likely to be considered as more significant than a neuron with only a few outgoing connections or low firing variance, and if the outgoing connections of the neuron are weighted strongly by other neurons then the significance of the neuron to the network is consequently higher. The described algorithm therefore differentiates between neurons which contribute significant variance to only a few other neurons (or to many neurons but with low weighting); those which are densely-connected with strong weighting but which contribute little variance; and those which are densely and strongly connected and which contribute significant variance. In other words, we can identify those neurons which simply "go with the flow" of the network rather than informing substantial areas of the network of a recent change of state elsewhere.

4.3.3 Information contribution of a single neuron

The significance measure outlined in section 4.3.2 gives an indicative measure of the contribution of the neuron to the wider network. However, this would be very hard to estimate biologically, as it requires precise knowledge of the synaptic weights of each of a neuron's afferents.

A more objective measure, often used in computational neuroscience due to its relative ease of calculation, relates to the information (in bits) provided by a neuron in terms

of its specific response to each of a set of stimuli. The information contribution of a single neuron can be measured as the *mutual information* between a random stimulus (for example, the requested retrieval one of the stored patterns) and the particular firing response this elicits from the neuron. The reduction in uncertainty of the identity of the specific stimulus, given by the response of the neuron, is a measure of the neuron's information contribution.

Skaggs et al. (1993) define mutual information between a neuron and a stimulus as "the information conveyed by a discrete random variable X about another discrete random variable Y , which is identical to the mutual information of X and Y". The same measure is used by Rolls et al. (1997) and Borst and Theunissen (1999) in the context of neural stimuli and responses, but with the notation s (stimulus) and r (response) instead of X and Y. The measure used is:

$$I(R,S) = \sum_{i,j} p(s_j)p(r_i|s_j)log_2\frac{p(r_i|s_j)}{p(r_j)} \tag{4.3}$$

where $I(R,S)$ is the total information in bits conveyed by a neuron's responses $r_i \in \{0,1\}$ to stimuli $s_j \in S$. Doya et al. (2007) helpfully define this as "the number of yes/no questions required to establish the value of the random variable" – in this case, each bit of information provided by a neuron about the state of the network answers one binary question about the network state. For example, if there are 16 discrete patterns stored in the network, it would take $log_2(16) = 4$ bits of information to discover the identity of any one pattern currently being presented to the network. A neuron which provides 3 bits of information is therefore able to reduce the uncertainty, or *entropy*, by $2^3 = 8$ patterns.

In this model, the probability of a response $p(r_i)$ from neuron i can be obtained for each neuron by repeatedly retrieving stored patterns from the network and recording the number of times neuron i fires, normalised by the number of patterns retrieved during the trial; $p(s_j)$ is simply the number of times each pattern s_j is retrieved during the trial, normalised by the total number of patterns retrieved; and $p(r_i|s_j)$ is observed by

noting each time neuron i responds to each pattern s_j, normalised by the total number of responses of neuron i to all patterns in set S.

4.4 Results

To investigate these predictions regarding information contribution, a series of experiments was performed in which randomly-generated networks were lesioned with either simulated scaling-driven pathology or by randomly selecting neurons for deletion. The resulting network performance at each step, in addition to the average information contribution of the selected neurons at the time of deletion, was recorded and plotted against the total proportion of deletion. A second set of experiments examined the change in pattern retrieval time (measured in number of update iteration steps until a stable pattern is obtained) as the networks were lesioned via scaling-driven and random deletion.

All experiments were undertaken using the Tsodyks and Feigel'Man (1988) model described in the previous chapter, with the following network parameters: network size $N = 1200$, connections per unit $K = 150$, firing threshold $\theta = 0.048$, noise $T = 0.005$, learning rate $\gamma = 0.025$, external input strength (learning mode) $e_l = 0.065$, external input strength (retrieval mode) $e_r = 0.035$, coding rate $p = 0.1$, neurons deleted per step $\Delta d = 0.01$. With the exception of network size and connectivity density, all the parameters were drawn directly from the findings of Ruppin and Reggia (1995a). The weight matrix was sparsely connected in a Gaussian manner, such that the probability of two neurons being connected increases with their proximity to each other. Results were averaged over ten runs, with the graphs showing mean and standard devation over the ten runs.

4.4.1 Random deletion

To investigate the effects on information contribution and network performance of deleting neurons using scaling-driven pathology, it is first necessary to form a baseline for compar-

isons. To achieve this, a network was created and then lesioned by deleting neurons and all their connections in random sequence until all neurons had been deleted. This was done by iteratively removing all incoming and outgoing connections (setting $W_{ij} = 0$ and $W_{ji} = 0$) from the weights matrix for each selected neuron N_i.

On each run, a set of 20 patterns was stored in the network according to the learning rule in equation 3.1. The network was then progressively lesioned at random in steps of Δd, or 12 neurons per step. Before the actual deletion in each step occurred, the significance and information contribution measures for each of the selected neurons were obtained according to the algorithms in sections 4.3.2 and 4.3.3, and the mean values over the 12 neurons were recorded. The connections were then deleted to remove these neurons from the network. A round of synaptic compensation was then performed for the remaining neurons as described in section 3.2.3.

After each iteration, the performance of the network was evaluated by obtaining the average overlap of the network (equation 3.2) when presented with each of the stored 20 patterns for retrieval. The results of this random deletion are shown alongside the average significance and information contribution of the deleted neurons at each lesioning step in figures 4.2, 4.3, and 4.4.

The results show a largely constant mean significance and information contribution for the deleted neurons at each Δd step with large error bars (figures 4.3 and 4.4), indicating that there is no specificity and high variability in information contribution during the deletion process. At around 80% deletion the information contribution starts to decline as the network becomes so highly disconnected that performance is at its minimum (figure 4.2), whereas the significance measure declines linearly with deletion, as significance is proportional to the number of afferents for each neuron. Increasingly, the remaining neurons are unable to receive sufficient input activation to fire, leading to a decline in their average contribution to the network immediately prior to deletion.

There is a highly linear decline in network performance until a period at around 60% deletion when the decline deviates more strongly but continues overall on a similar

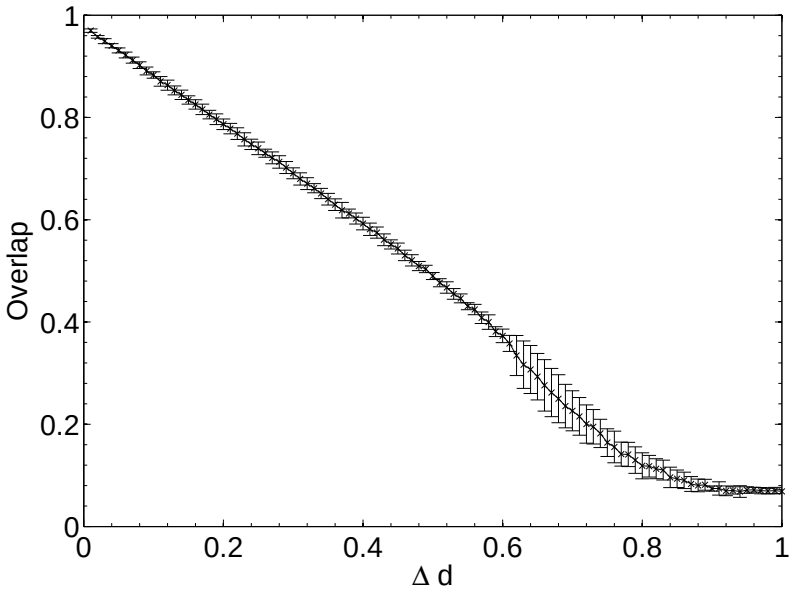


Figure 4.2: Network performance as neurons are deleted at random.

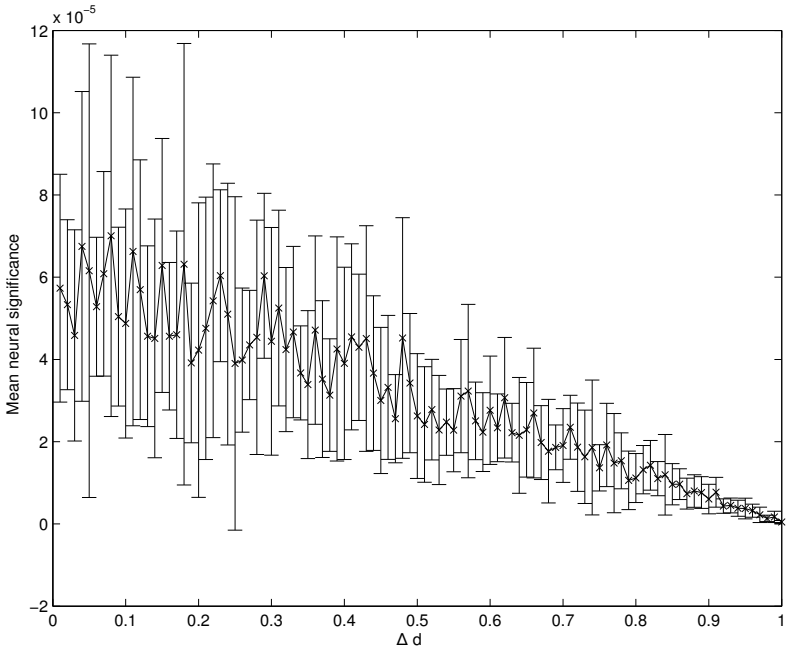


Figure 4.3: Mean significance Υ of neurons deleted at random.

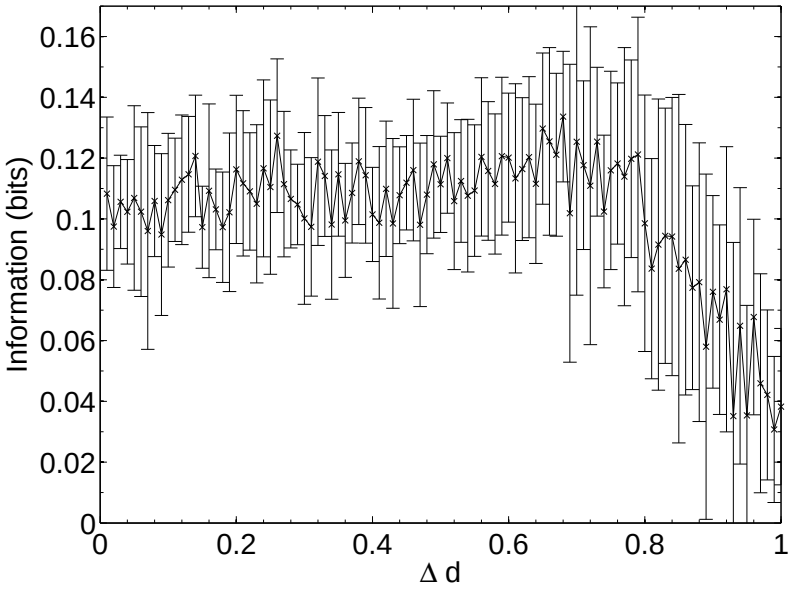


Figure 4.4: Mean information contribution of neurons deleted at random.

trajectory. The graph is clearly different to the s-shaped curve of performance decline seen in the previous chapter (compare figure 4.2 with figure 3.4), in which Δd of the remaining *synapses* chosen at random from across the whole network were deleted at each step, rather than deletion of all the synapses for the Δd selected *neurons* chosen at random.

Synaptic deletion grants each individual neuron a greater chance of remaining in the output pattern at each lesioning step, as it is highly unlikely that all the connections belonging to any one neuron are removed at each step of deletion, but with neuronal deletion the performance decline appears to represent solely the result of random removal of units from the output patterns, as each neuron in the network is compared with the stored patterns; there are no hidden units whose removal would have no effect on the output pattern. This is the equivalent of randomly flipping bits in the retrieved patterns to zero: the actual network effects of this deletion do not become apparent until the period of irregular performance decline from around 60% deletion onwards[1].

[1]The direct effects of neuronal deletion upon the output pattern (i.e. leaving aside any network effects) can be mitigated by separating the "processing" from the "output" in the network and lesioning only the processing layer, for example using a multi-layer perceptron or a reservoir network.

4.4.2 Scaling-driven lesioning

The experiment was then repeated with lesioning to simulate scaling-driven pathology according to the hypothesis of Small (2008). The neurons at each step were selected probabilistically as a linear function of the compensation factor c_i of each neuron. Again, 20 patterns were stored before the lesioning process was initiated, whilst results with differing numbers of stored patterns are analysed in section 4.4.4.

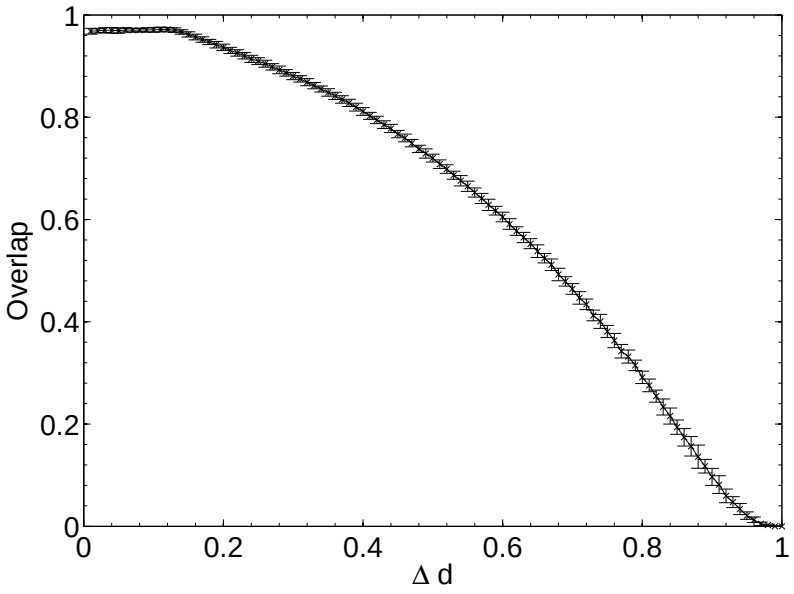


Figure 4.5: Network performance as neurons are deleted selectively using scaling-driven progression strategy.

These results show that, when considering significance and information contribution, the lesioning process initially acts at random within the network due to the flat distribution of the compensation factor c_i. The recorded significance and information contribution of the first set of deleted neurons is almost exactly at the mid-point between lowest and highest information contribution (figures 4.6 and 4.7). This random deletion also has a small negative effect on performance of the network, causing the overlap measure to begin just below 1 (figure 4.5).

Immediately after the onset of scaling-driven lesioning, as neurons with sparse and/or low-weighted incoming connectivity begin to lose their neighbours to the effects of Aβ,

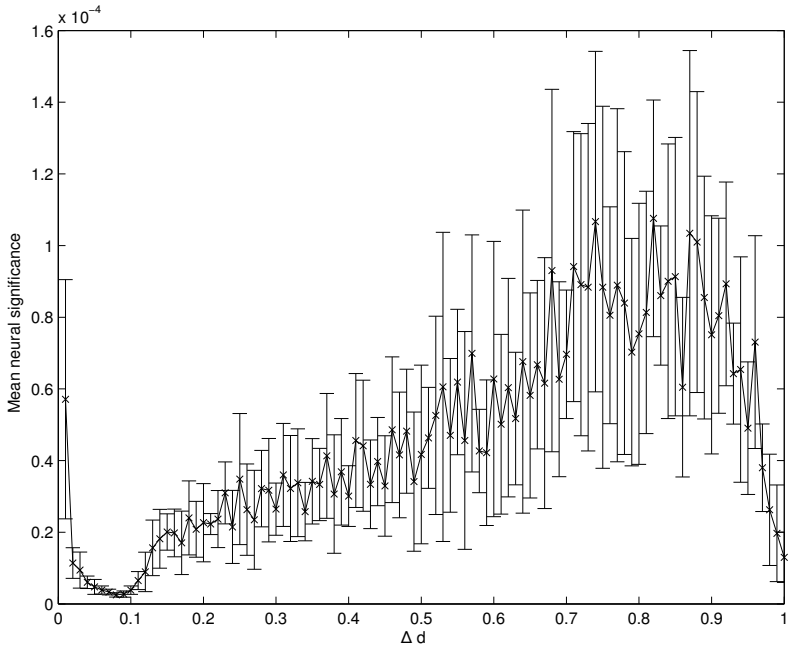


Figure 4.6: Mean significance Υ of neurons deleted selectively using scaling-driven progression strategy.

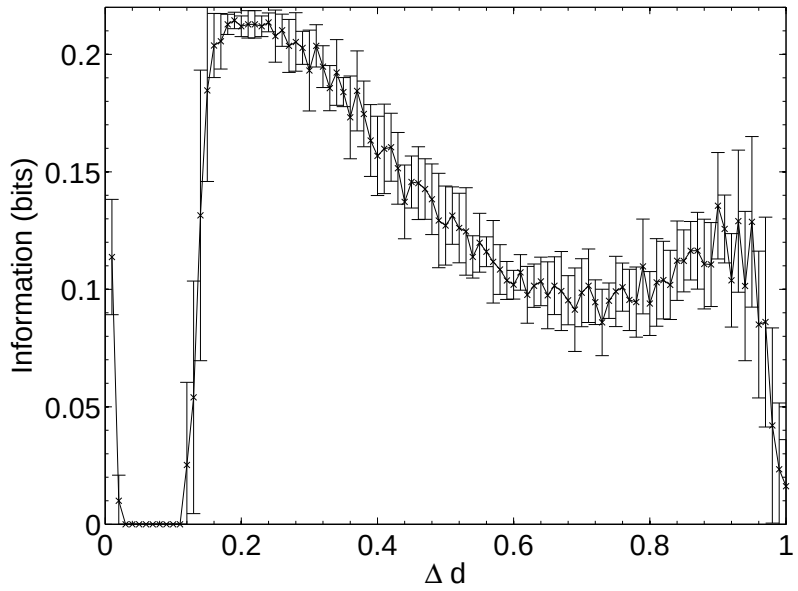


Figure 4.7: Mean information contribution of neurons deleted selectively using scaling-driven progression strategy.

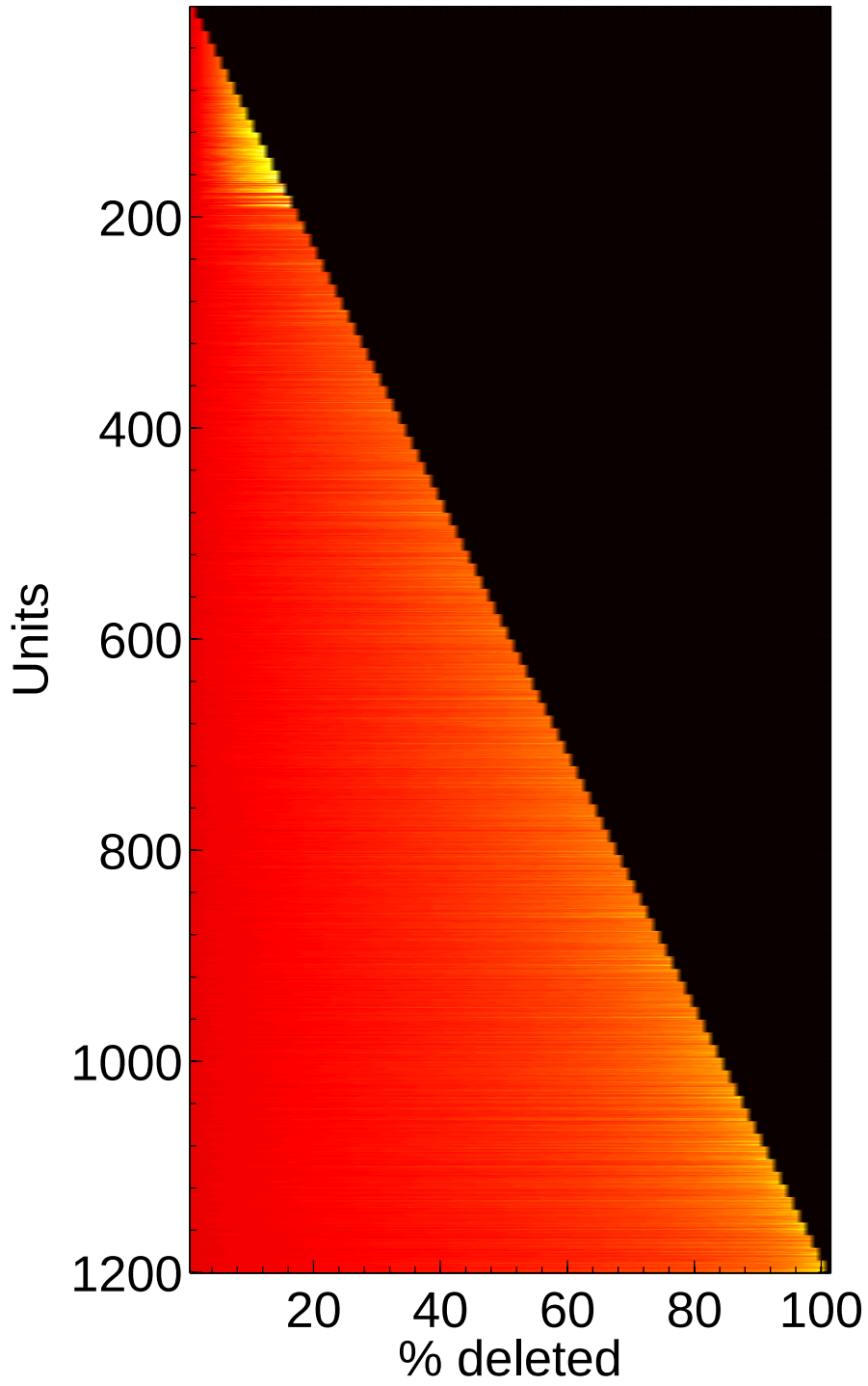


Figure 4.8: Change in compensation factor c_i as 1200 neurons are deleted, shown along the time course of the network from initial lesioning to 100% deletion. Hotter colours (lighter shades) denote higher values of c_i. Neurons nearer the top were deleted first; neurons on the bottom row survived all the way to the end.

their compensation rates must increase faster than those of better-connected neurons (figure 4.8). The pathology now selectively targets neurons with low overall contribution to the network due to the greater compensation rates they must employ. Because these neurons contribute relatively little to the overall functioning of the network and its ability to recall patterns (as suggested by their sparse and low-weighted connectivity), the performance does not decline at all for up to $12 - 13\%$ of deletion.

Once this reserve of low-significance / low-information contribution neurons has been removed, the significance of the neurons being selected for deletion begins to increase gradually, with the highest-significance neurons (that is, the neurons with the largest variance over the set of stored patterns and the strongest afferent connectivity) spared until the end of the process (figure 4.6). At this stage deletion of the neurons begins to have a tangible effect on the performance of the network (figure 4.5).

The information contribution also increases after the initial cognitive reserve is exhausted, but it increases much more rapidly to its peak at around $\Delta d = 0.2$ before declining with ongoing deletion (figure 4.7). This disparity between neural significance and information contribution could be explained by the decreasing performance of the network as a whole during this process (figure 4.5): whilst the neurons are selected for deletion increasingly in proportion to their afferent connectivity strength, the actual ability of those neurons to affect the remainder of the network decreases as the network becomes more disconnected, thus lowering their effective information contribution.

This result could explain the observed poor correlation between Aβ plaque density and cognitive decline in early AD (Hardy and Selkoe, 2002; Savioz et al., 2009; Minati et al., 2009), at least in the case that levels of amyloid are higher than expected for a given (low) amount of cognitive decline. During the first $12 - 13\%$ of neural atrophy due to scaling-driven progression in the model with 20 patterns stored, only insignificant (in information contribution terms) neurons are removed and there is correspondingly no apparent cognitive decline whatsoever. It is only after this point, as the pathology continues to selectively target the least significant of the remaining neurons, that cognitive

decline becomes apparent as the average information contribution of the dying neurons, and hence the negative effect on the network of their removal, gradually increases.

The trajectory of the significance and contribution of the remaining neurons then continues steadily between $20\% - 70\%$ deletion as the network progressively becomes more disconnected until, towards the end of the lesioning process, there is a second, smaller peak in information contribution which occurs as the last neurons to be deleted are the most highly-connected, significant neurons which are responsible for the above-zero recall performance even at this late stage. Finally, at around 90% deletion, the significance and information contribution of the deleted neurons decreases rapidly to nearly zero as the small numbers of remaining neurons are finally disconnected from each other, meaning there is essentially no "network" left for the neurons to contribute to.

4.4.3 Retrieval times

The experiment was repeated once more for both scaling-driven and random neuronal deletion, but instead of examining the decline in performance over Δd, the average number of network update iterations required to fall into a stable state for each cued pattern was recorded. This was done in order to discover whether the pathology has any effects on the stability of the network, and whether it affects the network's ability to fall into associative recall states.

Figure 4.9 shows that, for random neuronal deletion, the length of time for pattern retrieval increases steadily until the peak (set artificially at $0.05N = 60$ update iterations to prevent infinite updating which never converges) is reached. This is to be expected, as the removal of neurons and their connections at random from the network will necessitate longer travel paths of activation between the remaining neurons, and hence longer overall network update time.

The results of scaling-driven lesioning on average pattern retrieval time (figure 4.10) show a similar performance at the start of the lesioning process to that obtained during random deletion, but unlike in random deletion, the retrieval time actually begins to

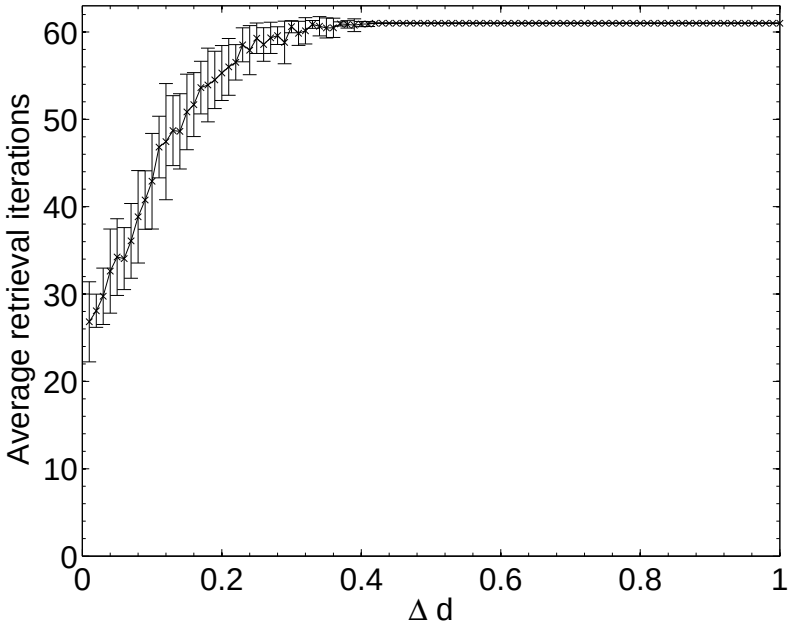


Figure 4.9: Changes in network pattern retrieval time during random neuronal deletion.

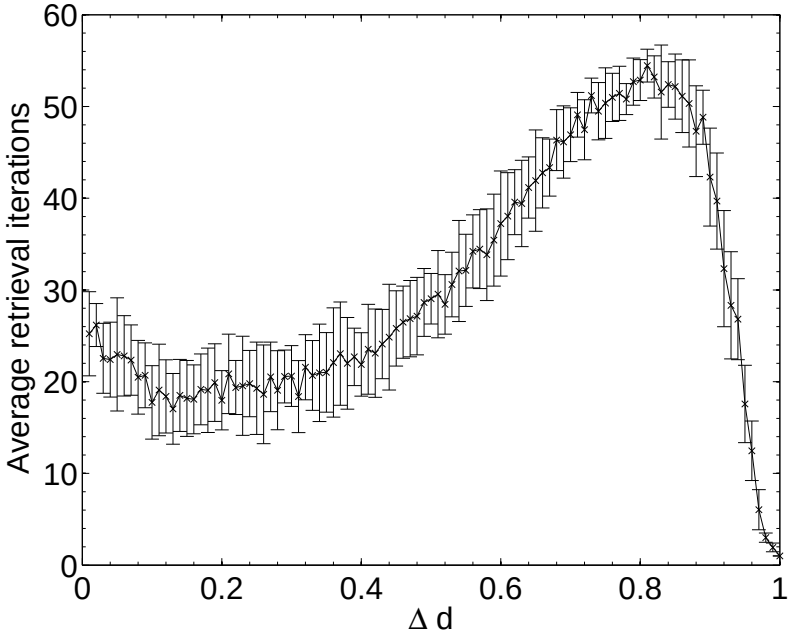


Figure 4.10: Changes in network pattern retrieval time during scaling-driven lesioning.

decrease over the first 10% of lesioning (the same period during which network performance remained stable and information contribution of deleted neurons was at its lowest in figure 4.5). This implies that the network performance is actually being optimised by the removal of low-contribution neurons, although (due to the effects of neuronal deletion mentioned in section 4.4.1) these early deleted neurons are likely to be ones which do not fire in any of the stored patterns and hence have very low variance and correspondingly low information contribution, as their removal appears to have no associated negative impact on the overlap measure.

Once the performance effects of neuronal deletion begin to appear at around $12-13\%$ deletion, the retrieval time steadily increases. This appears intuitive if it is considered that, as the deletion progresses, only the most significant neurons (in terms of operation of the network) remain: it is only once these neurons are being deleted, at around 80% deletion onwards, that both the information contribution of the deleted neurons and the retrieval time of the network reaches its peak. After the peak, at around 90% deletion, the network has become so fundamentally disconnected that these small numbers of highly-significant neurons finally become disconnected from each other, leading to a network which does not respond to the input cues at all but just remains in a stable state unrepresentative any of the cued patterns.

4.4.4 Network loading levels

To investigate potential effects of the loading level of the network on the results (in other words, to determine whether the results obtained are dependent on the number of patterns stored as a proportion of the network's maximum capacity), the experiment was repeated with increasing sizes of pattern sets.

The theoretical capacity of an associative network is $\frac{N}{2logN}$ patterns (McEliece et al., 1987) (where N is the number of units in the network), which gives a theoretical capacity for a 1200 unit network of 194 patterns, although as the network is connected sparsely ($K = 150$ connections per unit) the actual capacity is significantly lower. To find the

limit of the network's capacity experimentally, patterns were sequentially stored and later retrieved, and performance according to the overlap measure (equation 3.2) was obtained. The results in figure 4.11 show that the average maximum capacity over 50 trials for a network with this particular size and connectivity, beyond which performance declines significantly, is around 115 patterns, but that performance also declines linearly prior to this capacity limit as a function of the number of patterns stored. The error bars begin to expand at around 90 patterns, indicating that some networks had already started to fail when they reached this number of stored patterns. Therefore, over the 50 trials analysed, the maximum *safe* capacity (below which no network was seen to fail) was deduced to be approximately 90 patterns.

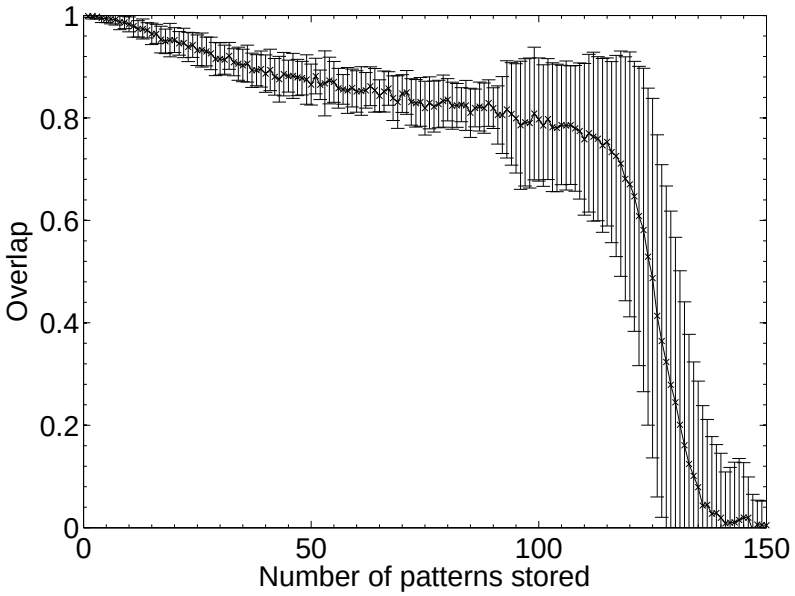


Figure 4.11: Network capacity (in number of discrete patterns stored) over 50 networks with $N = 1200$ units and $K = 150$ Gaussian connections per unit.

Results were then obtained for experiments in which 10, 15, 25, 30, 40 and 50 patterns respectively were stored in different networks, which were then lesioned by the same scaling-driven process outlined previously. The results for network performance and information contribution for these varying network loading levels are shown in figures 4.12 and 4.13.

These results demonstrate that the extent of the information-selectivity effect depends

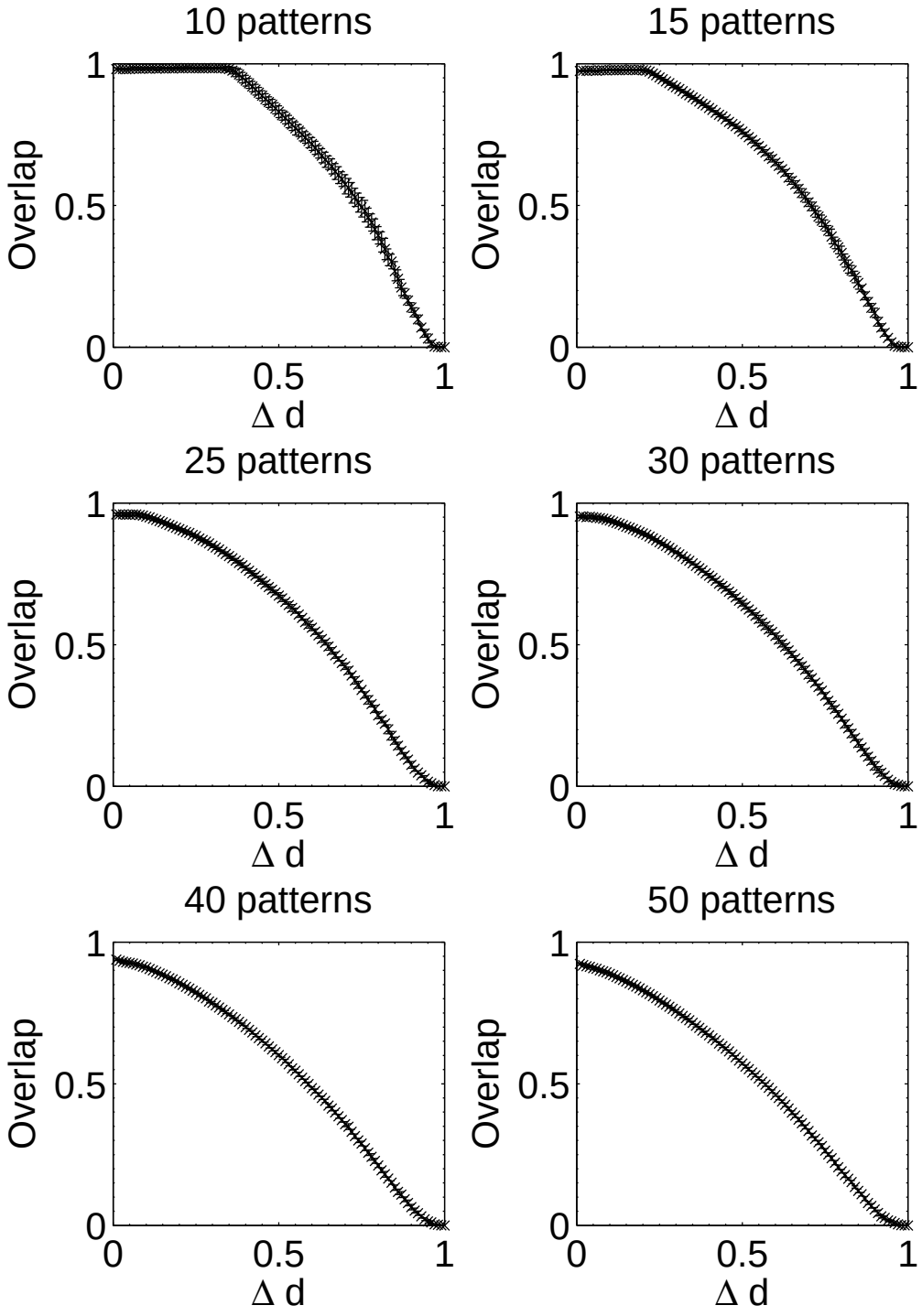


Figure 4.12: Network performance vs Δd as neurons are deleted selectively using scaling-driven progression strategy, for various sizes of pattern sets.

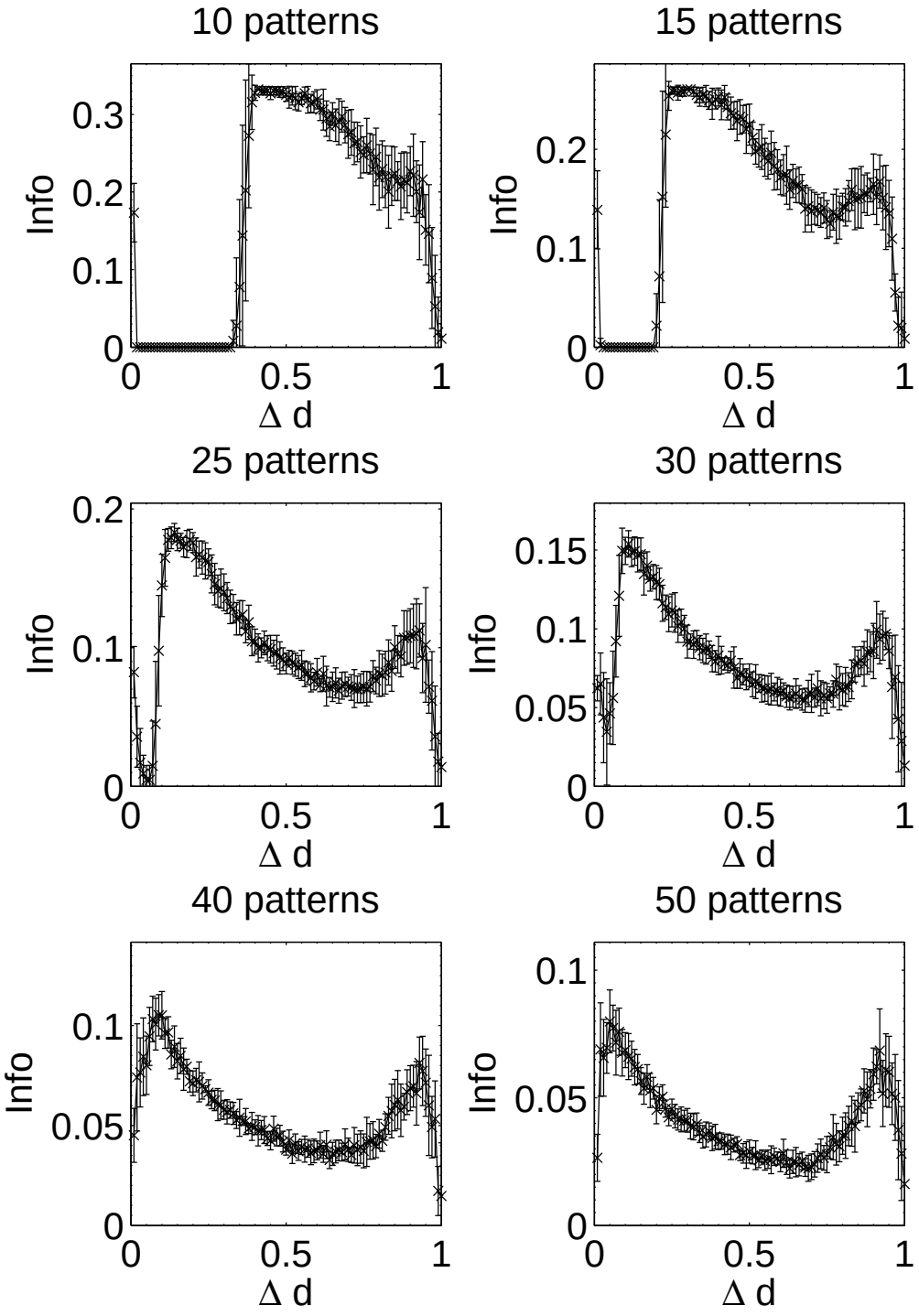


Figure 4.13: Mean information contribution of neurons deleted selectively using scaling-driven progression strategy, vs Δd, for various sizes of pattern sets.

on the loading level of the network. At low to medium loadings such as 10, 15, 20 (figure 4.7), and 25 patterns per network, the effect is clearly visible albeit reduced as a function of the loading level. Beyond 25 patterns per network the effect becomes less pronounced, before all-but-disappearing at 50 patterns (which is more than a 50% loading level when considering that the maximum safe capacity was found to be less than 100 patterns). Even at this high loading level, the initial few deletions are still below the peak information contribution per neuron before the general information decline begins.

At these higher loading levels (40 patterns and above), the network is already noticeably below optimal retrieval performance even before lesioning begins, as can be seen in figure 4.11, and this could account for the disappearance of the information-selectivity effect. Loading at these high levels may well be unreasonable in the biological context of the brain due to the increased noise, overlap between patterns, and non-optimal retrieval performance (Treves and Rolls, 1991), meaning that the principle outlined here should still be valid in the context of brain-like loading levels.

4.5 Discussion

This chapter has introduced a method for lesioning of computational neural network models via simulated beta-amyloid related pathology according to Small's hypothesis of scaling-driven synaptic progression (Small, 2008) in an associative memory model incorporating local, neural field-dependent synaptic compensation (Horn et al., 1996).

Whilst Small highlights the association between high-activity, highly-connected neurons (e.g. cholinergic and glutamatergic pyramidal cells) and vulnerability to beta-amyloid pathology, existing knowledge does not explain the observation that levels of amyloid deposits do not correlate well with cognitive deficits in early stages of AD (Hardy and Selkoe, 2002; Savioz et al., 2009; Minati et al., 2009). This work presents useful insights into the behavioural characteristics of Aβ-related scaling-driven pathology within an associative neural network, and provides a potential answer to that question using information

theory to measure empirically the information contribution of individual neurons to the network before they are lesioned. It has been shown that scaling-driven progression of AD pathology could selectively target neurons with low information contribution following the onset of AD with initially no adverse effect on network performance, but as the reserve of neurons with low information contribution is exhausted, the pathology begins to target neurons with increasing information contribution to the network, and the cognitive decline becomes more apparent and accelerative.

Interestingly, the model also suggests that rather than adversely affecting cognitive performance in the early stages of AD, the preference of scaling-driven pathology for neurons with low information contribution may actually lead to a small network optimisation effect via a form of neural Darwinism, as the removal of relatively insignificant neurons from the network results in shorter pattern retrieval times as well as a small reduction in energy requirements due to the reduction in neural volume. However, this effect is likely to be so small in a real brain as to be very difficult to measure experimentally, and may indeed be insignificant.

This type of neural network model is an abstract representation of biological neural networks, making many assumptions regarding the low-level neurobiological processes occurring in Alzheimer's disease, and it is also far from a complete model of cortical function, but care has been taken to ensure biological plausibility wherever approximations have been made (e.g. sparse neural coding in stored patterns (Abeles et al., 1990), sparse connectivity between neurons (Churchland and Sejnowski, 1994), activity-dependent rather than one-shot learning (Ruppin and Reggia, 1995a), local field-dependent synaptic compensation (Horn et al., 1996)). The model can be shown to fit into the general context of cortical-hippocampal learning, with the network designed such that it obeys O'Reilly and Rudy's principle that "the neocortex slowly [i.e. over extended periods of pattern presentation] learns about the general statistical structure of the environment, using overlapping distributed representations" (O'Reilly and Rudy, 2000), and incorporates backprojections from the hippocampal CA3 to the cortex as according to Rolls (2010); Rolls and Kesner

(2006), embodied here in the external input connections to the model which specify enforced activity of units during learning.

Nevertheless, the principles of the relationship between connectivity density and compensation rates explored in section 4.3.1 and demonstrated in the model are sufficiently general that they should apply to any neural network (artificial or biological) in which a scaling-driven pathology such as beta-amyloid targets neurons as a function of their increased activity as part of a compensatory mechanism when other connected neurons are targeted by the pathology. It is noted that biases appear when applying information theoretic measures to data, particularly with limited sample sets, and therefore the "true" information contribution measure of neurons in this work are subject to such biases (Paninski, 2003). This is an issue which could be addressed in future work, perhaps by applying Paninski's methods, or those of Panzeri and Treves (1996).

This work has also demonstrated that neuronal deletion in an associative network model suffers from the fundamental problem that, as all neurons in an associative network are the "output layer", the deletion of neurons from the network directly alters the output patterns, thus artificially reducing its performance even though the network may not actually be processing its input cues any differently. Such effects could be mitigated in future experiments by functionally separating "processing" from "output" in the model, for example by using a reservoir network (a recurrent neural network with a separate, trained, linear readout layer) and lesioning only the reservoir portion whilst the readout layer remains intact (Lukosevicius and Jaeger, 2009).

Finally, Alzheimer's disease consists of multiple interacting pathologies including beta-amyloid progression and the hyperphosphorylation of tau protein into neurofibrillary tangles. The previous chapter introduced a method for simulating tau pathology in the Ruppin and Reggia model of AD, and combination of this tau pathology simulation with the beta-amyloid-related scaling-driven pathology explored in this work could lead to further insights into the behavioural characteristics of the disease mechanisms. In particular, the precise relationship between the various pathologies is still poorly understood, but is

likely to consist of a number of complex self-reinforcing cascades of degradation.

4.6 Chapter summary

In this chapter, a hypothesis concerning the *information-selectivity* of amyloid-related scaling-driven pathology in Alzheimer's disease has been presented. This hypothesis describes how, in accordance with the progression hypothesis of Alzheimer's disease (Small, 2008), the pathology is predicted to select initially those neurons with low information contribution to the network. Thus, the 'cognitive reserve' is selectively targeted first as the disease spreads to multiple regions of the brain, with little or no apparent cognitive decline, therefore hiding the presence of the disease and making early intervention and treatment more difficult. These predictions could also explain at least one aspect of the disparity between numbers of amyloid plaques in the brains of AD patients and levels of cognitive decline.

The hypothesis outlined in this chapter, with the associated computational model, and subsequent predictions of information-selectivity in Alzheimer's disease are new, and have not been described before in the literature.

CHAPTER 5

BIOLOGICALLY-REALISTIC TESTING OF PREDICTIONS

Chapter synopsis

Previously, we introduced the information-selectivity hypothesis of synaptic scaling-driven pathology in Alzheimer's disease using a basic associative Hopfield-class model. This model involves many abstractions, which makes demonstration of the basic principles relatively straightforward, but also removes a lot of the detail required to make biologically-testable predictions. By exploring the information-selectivity principle in a biologically-realistic spiking model which exhibits emergent brain-like behaviour, it may be possible to look for potential bio-markers of the disease before the onset of cognitive symptoms, to enable earlier intervention and more successful treatment. Such a realistic model may also provide the opportunity to experiment with novel treatments which could not be attempted in an over-simplified model.

This chapter gives a background to the biological mechanisms of synaptic scaling. It then introduces the Yale ModelDB, a repository of peer-reviewed computational models, from which we evaluate a selection of candidate models in order to ascertain their suitability for modelling synaptic compensatory mechanisms and exploring the information-selectivity hypothesis. Consideration is given to the open problem of setting a neuron's target activity level, which is a necessary component of synaptic scaling. Due to the more

complex dynamics of networks in comparison to a basic associative network model, it is necessary to modify the information-contribution calculations to account for spiking neurons. Therefore, an overview is given of a fast method, utilising Fourier transforms, for estimating the information conveyed by a population of spiking neurons (Crumiller et al., 2011).

5.1 Biological mechanisms of synaptic scaling

Basic associative neural network models have their place in helping to clarify or simplify the effects of certain processes (such as synaptic scaling) on neural dynamics. But to be able to test these processes in more realistic scenarios, we must turn to biologically-realistic spiking models of brain regions; preferably those which have been validated according to existing knowledge of the brain. Such models allow not only more rigorous and detailed testing of hypotheses, but also the option to observe changes in network behaviour which should map onto biological predictions, thus enabling *in vivo* or *in vitro* testing of hypotheses, and perhaps earlier prediction of disease presence than is currently available through observing cognitive symptoms.

Due to the immensely time-consuming process of creating such a biologically-realistic model and obtaining experimental data with which to validate it, it was proposed instead to search for existing models which fit these criteria, and to extend one to include the synaptic scaling and corresponding toxicity mechanisms.

When choosing a candidate model to be extended, it is necessary to consider precisely which aspects of the synaptic scaling mechanism will need to be modelled. Small (2008) presents the following mechanisms through which compensatory synaptic scaling is believed to operate (refer to the glossary in section 2.1.1 for a definition of terms):

Stage 1: Toxicity Aβ induces loss of AMPA receptors in the dendrites of a neuron. This results in the neuron becoming less easily excited by the release of glutamate from presynaptic neurons, and therefore its firing rate decreases. At more advanced

stages of toxicity, Aβ can disrupt calcium homeostasis, leading to Ca^{2+} influx and the activation of cell-death pathways, causing the neuron to stop firing altogether and die.

Stage 2: Scaling Postsynaptic neurons receive less activation as a result of the reduced activity (or death) of presynaptic neurons in stage 1. Without scaling, this would lead to an overall decrease in network activity, and ultimately the cessation of all neural firing. The postsynaptic neurons must therefore scale up their activity levels via the following mechanisms:

Postsynaptic AMPA receptor increase The reduction in postsynaptic activation results in an increase in postsynaptic AMPA receptors in the dendrites of the affected neurons, causing these neurons to become more easily excited by the neurotransmitter glutamate, and resulting in restored firing rates and maintained average activation across the network (Turrigiano, 2008; Small, 2008).

Postsynaptic BDNF decrease Cortical pyramidal neurons release neurotrophic BDNF as a function of their average levels of activity. As the activity of a postsynaptic neuron decreases, extracellular BDNF levels also decrease. This leads to an increase in signalling strength amongst surrounding excitatory pyramidal neurons (conversely, an increase in BDNF would lead to an increase in firing of inhibitory interneurons, thus reducing overall activity), and the average activation is restored (Turrigiano, 2008; Small, 2008).

TNF-α increase Reduced neuronal activity causes *glial cells* within the brain to release higher levels of extracellular TNF-α, which leads to an increase in the number of AMPA receptors in the dendrites of surrounding neurons, a corresponding increase in excitability, and the restoration of average activity to the network (Turrigiano, 2008; Small, 2008).

ACh increase The decrease in activation of the postsynaptic neurons also results

in an increase in the release of acetylcholine by cholinergic neurons in the basal forebrain, leading to greater activation of the postsynaptic neurons. Small (2008) highlights the observation that "although a decrease in cholinergic activity in the brain is viewed generally as contributing to cognitive decline in AD, cholinergic activity might be increased during the very early (preclinical) stages of the disease", and indeed, an increase in cholinergic neuron activity could lead to excitotoxicity and their eventual death as the disease progresses (see stage 3).

Stage 3: Spread of pathology Compensatory scaling of postsynaptic neurons leads to an increase in their excitation, meaning that they must fire more strongly and/or frequently to maintain the average activation levels across the network. This results in a corresponding rise in intracellular Ca^{2+} levels within the remaining neurons, which makes them more susceptible to Aβ toxicity and the initiation of cell-death pathways via the disruption of calcium homeostasis, and the cycle recommences at stage 1.

Mechanisms of synaptic scaling are described in much greater detail by Turrigiano (2008). Importantly, it is notable that whilst long-term potentiation acts very quickly to strengthen synapses during Hebbian learning, synaptic scaling in neocortical and hippocampal pyramidal neurons is a relatively slow process (on the order of hours). It is believed that "neurons can detect changes in their own firing rates [from some target rate] through a set of calcium-dependent sensors [as calcium levels in the soma reduce with lower excitation] that then regulate receptor trafficking to increase or decrease the accumulation of glutamate receptors [AMPA or NMDA] at synaptic sites". The precise mechanism by which excitatory AMPA and NMDA receptors accumulate, however, remains unknown (Turrigiano, 2008).

The implementation of synaptic scaling in the selected computational model could be achieved in two ways: either by incorporating all of the detailed neurotransmitter-related

mechanisms identified by Small (2008) and Turrigiano (2008) to create a realistic biophysical model of the process, or alternatively by ignoring the specifics of the mechanisms involved and simply simulating the end results of the process (i.e. artificially increasing activity in postsynaptic neurons after a decrease in average activation from presynaptic neurons).

If synaptic scaling were to be modelled to a realistically high degree of accuracy, then every element of the mechanism and all the associated neurotransmitters, neurotrophins, and receptors as described by Small (2008) and Turrigiano (2008) would have to be present within the candidate model. Simulation environments such as NEURON offer the ability to add extra receptors and channels to the simulated neurons to simulate every possible element of detail (for example, Morse et al. (2010) studies the abnormal excitability of specific parts of a single neuron's dendrites in early AD), but implementing synaptic scaling at such a level of detail would add a large number of extra variables to the model which must be tuned and tested in order to reach a biologically valid "working state".

Alternatively, an approach similar to that of Cutsuridis et al. (2010) could be undertaken. In this work, the authors took the established principle of Spike Timing Dependent Plasticity (STDP, van Rossum et al. (2000)) and implemented STDP-driven encoding and retrieval of patterns in the CA1 area of the hippocampus. However, they did not attempt to model the detailed molecular mechanisms underlying synaptic plasticity in learning, but instead used a mathematical learning rule to increase the strengths of the synapses between the respective neurons. In the same way, the implementation of synaptic scaling could be achieved by monitoring average firing rates of each neuron and artificially increasing the strength of incoming synapses along a smooth profile according to the timescales identified by Turrigiano (2008). van Rossum et al. (2000) present such a mathematical rule. Another implementational approach (if the model provides such a facility) would be to change the number of simulated AMPA receptors in the affected neuron in proportion to the amount of scaling required, leading to discrete rather than continuous weight changes.

5.2 Candidate models

5.2.1 ModelDB

The Yale ModelDB[1] (Hines et al., 2004) presents a comprehensive repository of peer-reviewed computational models with accompanying source code covering various areas of neuroscience, including abstract connectionist models of neurological disorders such as Parkinson's disease, highly detailed multi-compartment models of individual neurons with separately-modelled soma, axon and dendrites, and networks of biologically-realistic neurons. Some of the models, such as Hasselmo and Eichenbaum (2005), perform some specific applied task such as simulating context-dependent retrieval relating to the movement of a rat in a virtual world, whilst other models focus on the firing patterns of the neurons themselves in order to visualise the emergence of various different firing rhythms (Bhattacharya et al., 2011; Bartos et al., 2002; Tort et al., 2007; Kopell et al., 2010; Stacey et al., 2009; Traub et al., 2005; Neymotin et al., 2011c).

5.2.2 Search requirements

To obtain a list of candidate models for extension to model syaptic scaling-driven pathology and information-selectivity, the ModelDB was searched for potentially useful publications on the basis of the following requirements:

- The model should simulate a network of neurons (as opposed to a highly-accurate model of a single neuron, such as Morse et al. (2010)).
- The model should include recent experimentally-derived biophysical connectivity data appropriate to the hippocampus or the neocortex (two brain regions highly implicated in Alzheimer's disease, and for which large volumes of experimental data are available).

[1] http://senselab.med.yale.edu/modeldb/

- The model should include as many as possible of the neurotransmitters, ion channels, and neurotropins believed to be responsible for synaptic scaling and associated excitotoxicity from those listed by Small (2008) (AMPA, NMDA, ACh, BDNF, TNF-α, glutamate, and Ca^{2+}).

- The model should have been experimentally validated, whether by the authors themselves *in vivo* or *in vitro*, or against available data from the literature.

- The model should perform some task relevant to symptoms of Alzheimer's disease – for example, producing stable oscillations known to be affected in AD (Bhattacharya et al., 2011), or pattern storage and retrieval, so that alterations in these processes can be investigated as possible biomarkers for AD in clinical settings.

- The model should operate fast enough on available serial or parallel computers (if the model has been written to operate in parallel) to be able to obtain timely results for long runs (potentially hours or days of simulated time, in order to simulate synaptic scaling) at sufficiently large network sizes to be able to examine the information-selectivity effects.

- The model should be written in a suitable programming language for easy extension and fast operation, or for a widely-used simulation environment such as NEURON or PyNN, and all source code must be provided.

A number of models were selected from the ModelDB which appeared, upon first examination, to conform broadly with these criteria. The majority of the models focussed on the task of replicating some part of the physiological oscillations (alpha/beta/gamma/theta rhythms) of synchronised neural firing found within the hippocampus or neocortex. Examples of such models and the brain regions which they model include Traub et al. (2005) (thalamus and cortex), Neymotin et al. (2011c) (neocortex), Stacey et al. (2009) (hippocampus), Kopell et al. (2010) (hippocampus), Vierling-Claassen et al. (2010) (neocortex), Bartos et al. (2002) (hippocampus), Anderson et al. (2011) (cortex), and Bhattacharya et al. (2011) (thalamus and cortex). Only two potentially suitable models which

were designed to perform an applied practical task could be found in the ModelDB: Hasselmo and Eichenbaum (2005) (hippocampus), in which a simulated rat traversing a maze is able to retrieve memories dependent on context, and Cutsuridis et al. (2010) (also hippocampus) which simulates STDP-driven storage and retrieval of patterns in a biologically realistic associative network, with a pattern overlap measure similar to that used by Ruppin and Reggia (1995a). The applied tasks investigated in these two models provide an additional measure of network performance during a lesioning process, which could assist with analysis of the network.

5.2.3 Evaluation of models

The identified models were then examined more closely to determine how suitable they would be for extending to include synaptic scaling and beta-amyloid related pathology.

The model of Anderson et al. (2011), based upon an earlier model by the same authors (Anderson et al., 2007), was designed to explore the generation of epileptiform activity in areas of abnormal synaptic wiring in the neocortex. It simulates a large network of 65,536 point neurons with AMPA and GABA receptors, in seven different excitatory and inhibitory classes, arranged into 4096 cortical minicolumns. The connectivity is based upon a previous visual cortex model, which the authors justify in relation to its applicability to other areas of the cortex, and it is able to produce biophysically realistic bursting and cessation behaviour. However, the bursting and cessation characteristics of the network are highly dependent upon the precise connectivity between neurons, and as the model focuses on prediction and evolution of epileptic seizures, the "normal" operation of the network is not well-defined and would therefore require significant parametric tuning to be usable for sufficiently long enough runs to model synaptic scaling.

Traub et al. (2005) present a highly complex model of a single column of neurons in the thalamocortical region. The model uses 3560 multi-compartment neurons in a variety of classes, each with AMPA, NMDA, $GABA_A$ and $GABA_B$ receptors, to generate persistent gamma oscillations which were evaluated experimentally by the authors in vitro in rat

neural cells, and in vivo in living rats. The model concentrates on the sleeping state via simulation of 'sleep spindles', and genesis of epileptiform bursting activity. Many of the other models in this survey cite this work, and it appears to be considered as something of a benchmark model, however Neymotin et al. (2011c) note that it only "shows a single peak at 30Hz" in gamma oscillations in the sleeping state, and in the epileptiform state the spectrum produced by the model "has additional high-frequency peaks" not expected in a normal awake state. They add that more recent models have now been shown to produce more biologically realistic approximations.

Stacey et al. (2009) created a small network of 100 multi-compartment hippocampal neurons (5 simulated compartments per neuron) to investigate the potential for random synaptic noise to initiate high frequency oscillations, in a similar fashion to Neymotin et al. (2011c). This model was adapted from a previous model by Tort et al. (2007) which studied the formation of gamma oscillations in the hippocampus, and was extended to include extra noise inputs and recurrent coupling. It provides inhibitory GABA and excitatory AMPA and NMDA receptors, and pays particularly close attention to the role of *oriens lacunosum-moleculare* interneurons (a specific category of inhibitory neurons). However, as in Traub et al. (2005) and other models, only gamma frequency (26-100Hz) and some theta (less than 10Hz) oscillations are observed, and this model is specifically aimed at producing "fast gamma" and higher-frequency (up to 200Hz) oscillations, which are more typical of epileptiform behaviour, meaning that the model's representation of sustained "normal" hippocampal firing characteristics is incomplete.

Kopell et al. (2010) offer a simple model of 100 point neurons, but with only two types of inhibitory interneurons and one type of excitatory neuron. The model displays gamma (to a more precise definition than Tort et al. (2007)) and theta oscillations which match physiological data, and includes AMPA and GABA receptors, but omits NMDA receptors. The authors justify the use of point neurons as follows: "the gamma rhythms discussed here are believed to be formed near the somata of the pyramidal cells, justifying a one-compartment model of those cells" and point out that "the aspects that are important

to the gammma rhythm can be captured in models as simple as integrate-and-fire" but the overall lack of biological detail makes this model less suitable for making biologically-testable predictions.

Vierling-Claassen et al. (2010) investigated neocortical oscillations and attempted to reproduce low-frequency rhythms in a model of 48 neocortex neurons. AMPA and $GABA_A$ receptors were included, but not NMDA, although multiple neural compartments were simulated in a biologically realistic manner according to previously-obtained biophysical data. Despite this biological realism, the authors found that their model was "not sufficient to reproduce lower-frequency 8Hz enhancement", and also "not sufficient to accurately simulate a recorded LFP [local field potential]", meaning that any biological predictions which could be made from this model are likely to be fairly limited.

A candidate model specifically designed to investigate Alzheimer's disease is that of Bhattacharya et al. (2011). The model demonstrates good fitting of the computer simulation to recently-obtained experimental data showing that, in AD, alpha rhythm oscillatory power shifts towards the slower end of the alpha spectrum. As with the majority of other models, the network doesn't attempt to perform any applied task, but instead the mode of operation is via random stimulation to observe the flow of activation throughout the network. However, rather than modelling individual single- or multi-compartment neurons, this model 'clumps' neurons together into a so-called "lumped neural mass model" to give average spike rates, rather than examining individual neurons. This approach is justified by the authors in order to "provide a mathematical definition for the behaviour of a neuronal population so densely packed that they may be assumed to be a continuum", but it makes the model unsuitable for the purposes of investigating the information-selectivity hypothesis as this requires simulation of individual neurons. Cao and Grossberg (2011) provide a method of "unlumping rate-based models that use the membrane equations of neurophysiology into models that use spiking neurons", which may be applicable, although this would require extra work. A further paper (Abuhassan et al., 2011) uses a network of individual Izhikevic neurons (Izhikevich, 2004) to undertake a similar experi-

ment to Bhattacharya et al. (2011), but source code for this is not provided in ModelDB. The Bhattacharya et al. (2011) model is written in Simulink, which can be analysed automatically in Matlab.

Bartos et al. (2002) present a model investigating gamma oscillations in inhibitory GABAergic interneurons in the hippocampus. The authors performed their own experimental verification of the morphology of the model, which was programmed with 200 single-compartment point neurons, each randomly connected to its 100 nearest neighbours in a flat ring, which is perhaps over-simplified. Nevertheless, the model was still able to replicate "highly coherent oscillations over a wide range of frequencies (20-110Hz)" including gamma rhythms nested within longer theta rhythms (Bartos et al., 2002). However, for the purposes of investigating the information-selectivity hypothesis, the complete lack of excitatory AMPA and NMDA neurons is a major omission.

The model of Hasselmo and Eichenbaum (2005) builds upon the principle, also studied by Cutsuridis et al. (2010), of theta rhythms separating encoding and retrieval in the hippocampus (Hasselmo et al., 2002). In this model, which is simulated in Matlab, a virtual rat retrieves stored patterns in a context-dependent manner depending on its location within a virtual world, and uses these to guide its movements as it searches for a food reward with the benefit of previously-learned experience regarding the locations of food. The model "was constrained to fit the elecrophysiological recordings of synaptic currents during the theta rhythm, as well as the spiking activity of hippocampal neurons" using available experimental data from the literature, and additionally models the interaction between hippocampus and prefrontal cortex (which is not seen in any of the other models). However individual neurons in the model have no specific receptors such as AMPA, NMDA, or GABA, meaning that any biological predictions which could be made from this model would be limited in their accuracy.

Neymotin et al. (2011c) present a large network simulating a section of neocortex containing 4230 point neurons across 13 classes of inhibitory and excitatory cells, with excitatory AMPA and NMDA as well as inhibitory $GABA_A$ receptors on each cell. The

network is arranged into 9 cortical columns, giving 470 neurons per column. Connectivity data is taken from a variety of recent sources, although some estimates have had to be made, and data is taken from various different (albeit related) species. The model displays a full range of oscillation frequencies under background noise stimulation conditions, and benefits from an extensive analysis of the network's firing dynamics via the technique of rewiring neurons to slowly alter the distribution and number of "hubs" within the network, as well as detailed information-theoretic analysis of the flow of data throughout the network (Neymotin et al., 2011a), for which source code is also available in the ModelDB.

The Neymotin et al. (2011c) model contains fixed synaptic weights between neurons, meaning that no associative learning of patterns is possible and initial information contributions for each neuron may be arbitrary, but another paper based on this model (Neymotin et al., 2011b) implements a STDP learning rule in the AMPA receptors of excitatory neurons to tune the oscillatory dynamics of the model. The authors predict that "homeostatic control mechanisms must balance learning at excitatory-to-excitatory and excitatory-to-inhibitory synapses" in order to avoid saturation of firing and transition to an epileptic state, matching the predictions of Turrigiano (2008) regarding the need for synaptic down-scaling during learning.

The model has been experimentally verified according to data taken from rat neocortex, and shown to match the physiological data with a high degree of accuracy. The authors highlight the fact that their model shows oscillations similar to those of a 'waking' brain state, in specific contrast to the model of Traub et al. (2005) which shows a smaller range of oscillatory frequency peaks. The model benefits greatly from nearly real-time processing speed using the NEURON environment (i.e one second of simulated time takes approximately one second of processor time), which makes the model suitable for sufficiently long runs to simulate synaptic scaling.

Another strong candidate model, this time of the CA1 region of the hippocampus, was presented by Cutsuridis et al. (2010). The model demonstrates associative storage and retrieval of patterns in the CA1 region of the hippocampus via a simulated STDP learning

process. In a similar approach to Ruppin and Reggia (1995a), the authors provide an "overlap" measure of recall performance as patterns are stored and retrieved using a noisy cue, thus allowing a quantifiable analysis of the network's performance on a standardised task. The model investigates a hypothesis by Hasselmo (Hasselmo et al., 2002) which predicts that theta rhythms dictate a mode switch between episodes of storage of new information and retrieval of old information.

STDP is not modelled biophysically at the molecular level, but a biologically-plausible localised mathematical learning rule defines the strengthening and weakening of relevant AMPA synapses based upon realistic simulations of the appropriate firing patterns within the network. In this particular model, only a subset of the synapses on the pyramidal (excitatory) neurons are modifiable during the STDP learning process, and none of the inhibitory synapses are modifiable.

In addition to the applied task of learning and retrieving patterns in an associative manner, the model provides realistic simulations of gamma and theta rhythms with verification according to data from the literature on anaesthetised rats, and the network wiring data is taken from a variety of experimentally-derived literature sources.

Neurons within the model are multi-compartment (13-17 separately modelled compartments), in comparison to the single-compartment point neurons of Neymotin et al. (2011c), and are presented in five distinct classes (one excitatory and four inhibitory). The model provides various Ca^{2+} and K^{+} channels with AMPA, NMDA, $GABA_A$ and $GABA_B$ receptors, offering one of the most realistic simulations of all of the studied models. This means that there is a trade-off, which is that only 100 cells are modelled due to the additional complexity of the simulation. Even with this small network size, the model takes a long time to process a single second of simulation time. It is written to run in parallel using the NEURON environment, which mitigates some of the slow computation speed, but despite this, using 20 CPU cores of the University of Birmingham's BlueBEAR computing cluster, a speed of approximately 0.01 times real-time was obtained. In comparison with approximately real-time running for the Neymotin et al. (2011c) model, this

would make simulations on the order of "hours to days" (as required by synaptic scaling) intractible.

Conclusions

Given the high level of detail across multiple excitatory and inhibitory neuron classes, with multiple neurotransmitters and a large network size, it was decided that the Neymotin et al. (2011c) model should be extended to add synaptic scaling, scaling-driven progression of Alzheimer's disease pathology, and investigation of information-selectivity of the pathology within the network. The Cutsuridis et al. (2010) model offers an even higher degree of complexity and detail, in addition to a very useful associative pattern storage-and-retrieval schema of operation, but a key benefit of the Neymotin et al. (2011c) model is the real-time running speed, which is approximiately 100 times faster than the model of Cutsuridis et al. (2010). A fast running speed will be essential for modelling long-term slow synaptic plasticity such as compensatory synaptic scaling, which works on timescales of hours to days of simulated time.

5.3 Setting target activity points

Synaptic scaling requires the setting of a target activity level for each neuron, towards which the neuron will continually scale its activity. Neural firing rates fundamentally affect the behaviour of the network as a whole, as imbalances in the firing rates of excitatory and inhibitory circuits can push the network to epileptiform hyperactivity and prevent information processing (Neymotin et al., 2011b), or cause total hypoactivity and loss of activation. So, clearly, it is important to set the activity targets of cells within the neuron with care, so as not to disrupt the normal behaviour of the network.

It is not currently known exactly how a neuron's activity target is determined biologically. The literature suggests two possible mechanisms: Either the neuron's firing pattern may be determined explicitly as part of its phenotype during development (i.e. encoded in

its genotype), in the same way as its afferent neurotransmitter type and firing threshold, and expressed by ion channel density at each relevant synapse; or the firing pattern may be an emergent property of the neuron's phenotype (i.e. not encoded in its genotype), with the combination of genetically-determined factors such as neurotransmitter type and firing threshold leading to a characteristic firing rate (Marder and Goaillard, 2006; Marder and Prinz, 2002).

Davis (2006) points out that many ion channels, which determine excitability and thus "natural" firing rate, are free parameters which may themselves be varied under homeostatic control, lending support to the second of these two mechanisms. Indeed, it is also possible that the activity targets may not even be fixed once initially set. If the equilibrium between the opposing forces upregulating and downregulating excitatory AMPA receptor accumulation during scaling can be modulated, this would enable shifting of the balance point of the equilibrium and thus altering the activity target (Turrigiano, 2008).

These considerations pose a problem for a computational implementation, as models such as Neymotin et al. (2011c) are finely-tuned with weights set to achieve a balance between over-excitation and over-inhibition. Arbitrarily setting a target firing rate for each neuron is liable to invoke unnecessary scaling, and may fundamentally affect network dynamics, as each cell is being forced away from its "natural" firing regime. For this reason, the proposed solution is instead to allow the model to run at its natural baseline oscillatory dynamics for a suitable length of time, and then to use the mean firing rate of each neuron to retrospectively set its target activity for scaling. In this way, the network should remain stable during long periods with scaling active, as little or no scaling is required for each cell due to the setting of target activities derived from each neuron's intrinsic activity.

5.4 Obtaining information contribution

The mutual information calculation strategy introduced in section 4.3.3 is not easily applicable to large networks of spiking neurons, as there can be no precise definition of whether or not a neuron is responding to a stimulus due to the presence of random spikes caused by noise, and the temporal nature of the response. For example, the existing information calculation can only determine the binary response from each neuron to a pattern once the network has reached its stable state, but in spiking networks each neuron may fire either before or after any other neuron with which its firing correlates. The temporal ordering of this firing response is important, as it defines which neurons are the "drivers" of activity within the network, and which are simply "followers".

Several methods exist for extracting information from small populations of spiking neurons. These operate by calculating the entropy present in the various temporal spiking patterns of each neuron, and finding the reduction of this entropy when in the presence of certain stimuli and considering the responses of other neurons in the population. It is important, when considering the total information contribution of a population of neurons, to distinguish between the *sum-total* information of the population and the *group* information. The group information may vary significantly from the sum-total information, due to the effects of *synergy*, whereby the co-ordinated response of a group of neurons provides more information than each provides individually, and *redundancy*, whereby multiple neurons deliver the same information to the network.

Such approaches are reviewed in Quiroga and Panzeri (2009), but they all suffer from the same problem of computational intractibility when scaled up to large neural population sizes, such as the hundreds or even thousands of neurons present within the Neymotin et al. (2011c) model, due to the multi-dimensionality inherent in considering the responses of each neuron to all other neurons in the population.

A recent method introduced by Yu et al. (2010) and summarised in Crumiller et al. (2011) addresses these issues by obtaining the neural spike responses to a set of unique and repeated stimuli. The repeated stimulus invokes a characteristic response for the

network, the entropy of which can be subtracted from the entropy of the response to each unique stimulus trial, in order to remove intrinsic population noise from the calculations. By reducing each neural spike train to a series of Delta functions, each neuron's response can be represented as points on a set of sine and cosine functions, thus enabling Fourier decomposition of the spike train into the frequency domain, and a vast reduction in the dimensionality of the calculation. Such calculations are (by way of their Fourier nature) obtained over a range of oscillatory population frequencies from 0 Hz up to a frequency at which the entropy of the responses to the repeat and unique stimuli is the same, meaning that the total information contribution for each neuron is the sum of its information contribution at each Fourier freqency.

This method is capable of quickly calculating the individual information contributions of hundreds of neurons in a population, as well as obtaining the difference between the sum-total and group information to enable a measure of the synergy or redundancy of the population.

5.5 Chapter summary

This chapter has taken a detailed look at the biological mechanisms underlying synaptic scaling, and deduced from these a set of requirements for investigating scaling-driven pathology of Alzheimer's disease in a biologically-realistic computational model. The Yale ModelDB was searched for suitable models, and a number of candidate models were systematically evaluated, resulting in the selection of the Neymotin et al. (2011c) model of emergent oscillations in the neocortex.

Consideration was then given to the open problem of setting target activity levels for synaptic scaling mechanisms without disrupting intrinsic network dynamics, and how this might be implemented computationally. Finally, a recent technique for estimating the information contribution of large populations of spiking neurons using Fourier analysis was outlined, in comparison to the simplified estimation measure used in previous chapters.

CHAPTER 6

INTERACTIONS BETWEEN SCALING AND LEARNING IN A SPIKING MODEL

Chapter synopsis

Chapter 4 demonstrated a predicted negative side-effect of synaptic scaling in neurons which attempt to compensate for lost activation during Alzheimer's disease progression, and the subsequent prediction that this causes Alzheimer's pathology to selectively target low-importance neurons in the cognitive reserve, thus delaying the onset of cognitive symptoms and masking the presence of the disease, making timely intervention and treatment more difficult.

However, providing medication to reduce or block the effects of synaptic scaling, as originally suggested by Small (2008), could also have severe consequences for patients. This is because synaptic scaling does not just up-regulate activity in response to decreases in the output of neighbouring cells: it also plays an important role during learning, where synapse potentiation during Hebbian learning must be down-regulated to avoid positive feedback cycles and runaway hyper-potentiation.

This chapter introduces the well-known Spike Timing-Dependent Plasticity (STDP) learning rule, and the problem of hyper-potentiation during learning. It investigates the effects of unconstrained STDP in the biologically-validated spiking neural network model by Neymotin et al. (2011c) identified in the previous chapter as it is trained to produce

frequency shifts in its oscillatory dynamics. Synaptic scaling according to the algorithm of van Rossum et al. (2000) is added to the model and amended to cope with a variable simulation timestep.

Experiments are performed to show that scaling can prolong activity during random neuronal deletion in the compensatory up-regulation case, but also that it is stable at baseline and does not interfere with normal oscillatory network behaviour, even at exceptionally long run times. Previous experiments have only shown scaling on the order of minutes to hours, but here we observe stable synaptic scaling over days of simulated time. The hyper-potentiating effects of unconstrained STDP are demonstrated, leading to epileptiform firing, but it is shown that synaptic scaling is capable of preventing this by down-regulating synapses during learning.

Finally, Fourier analysis is performed to produce power spectra plots showing that synaptic scaling does not simply "undo" the learned changes in network behaviour, but that scaling is capable of preserving learning.

Work derived from this chapter has previously been published as Rowan and Neymotin (2013).

6.1 Introduction

Spike Timing-Dependent Plasticity (STDP), a phenomenological learning rule in which synaptic potentiation and depression depend upon relative firing times (Dan and Poo, 2004; Zhang et al., 1998) to achieve Hebbian learning (van Rossum et al., 2000), has been used to learn oscillatory rhythms in neocortical models. In an existing biologically-realistic spiking model of neocortex (Neymotin et al., 2011c), applying excitatory to excitatory (E→E) STDP with a rhythmic training signal led to hyper-potentiation through positive feedback: strengthened synapses drove postsynaptic neurons to fire immediately, leading to further potentiation. This unbounded potentiation then pushed the network into synchronised epileptiform firing. Directly opposing E→E learning with equal excita-

tory to inhibitory (E→I) potentiation partially balanced this positive feedback. However, epileptiform behaviour still occurred with high-frequency signals (Neymotin et al., 2011b).

Homeostatic synaptic scaling is a local feedback mechanism which senses levels of activity-dependent cytosolic (intracellular) calcium within the cell and adjusts neuronal firing activity accordingly. This is achieved by producing alterations in excitatory AMPA receptor accumulation on the dendrites of the neuron, in response to changes in firing activity occurring over hours to days (Turrigiano, 2008), leading to changes in the excitability of the neuron.

During learning, synaptic scaling plays an important role in balancing potentiation. By constantly shifting mean activation back towards a target activity level, but maintaining the learned relative distribution of presynaptic weights, global levels of activity can be regulated (van Rossum et al., 2000). During periods of hypoactivity (for example in degenerative disorders such as Alzheimer's disease), synaptic scaling is also capable of raising the sensitivity of neurons via AMPA receptor upregulation, so that activity levels can be restored (Turrigiano, 2008).

Previous work has demonstrated synaptic scaling with learning in a single-neuron model (van Rossum et al., 2000). It has also been shown that synaptic scaling can prevent input saturation in a spiking neural network in the absence of learning (Chandler and Grossberg, 2012). Here, long-term synaptic plasticity is added to a spiking neural network to show that homeostatic synaptic scaling can prevent hyper-potentiation whilst preserving learned information.

6.2 Model description

6.2.1 Spiking model of neocortex

The spiking model used in this work was an extension of the neocortical model used in Neymotin et al. (2011b). It is based on the notion of a single cortical column (Neymotin

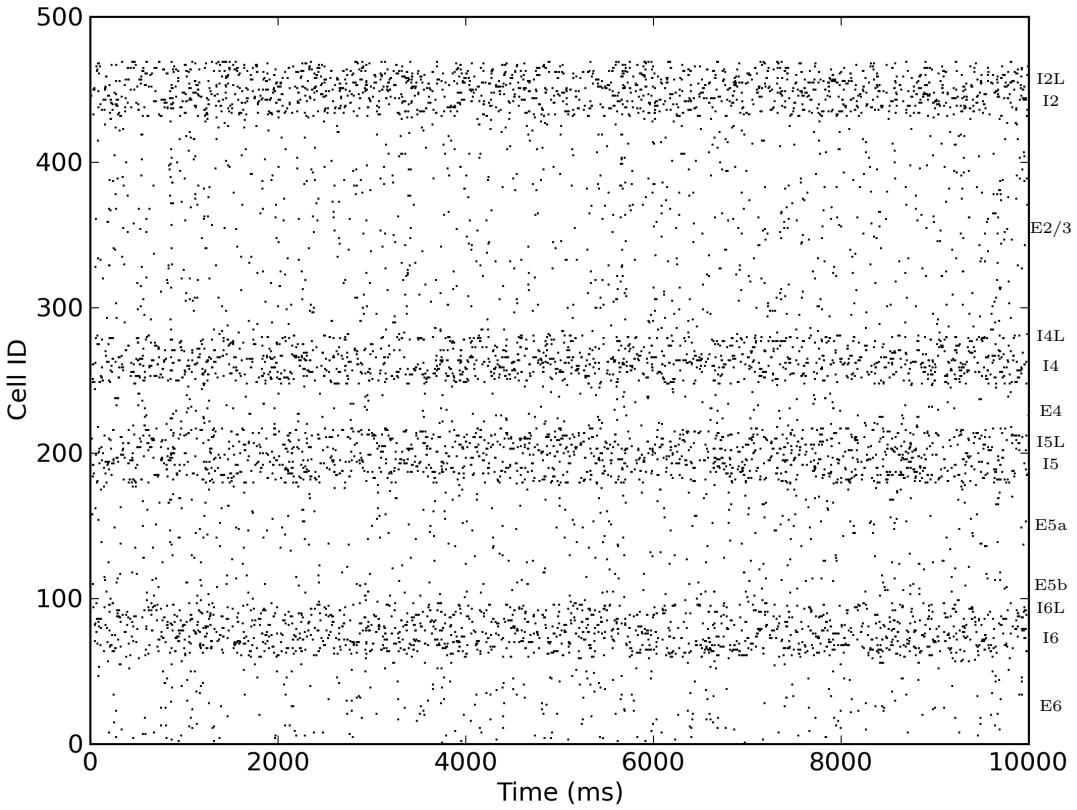


Figure 6.1: Raster plot showing distinct layers in the cortical column.

Layer	I2L	I2	E2	I4L	I4	E4	I5L	I5	E5a	E5b	I6L	I6	E6
Pop size	13	25	150	14	20	30	13	25	65	17	13	25	60

Table 6.1: Numbers of cell per type in each layer. E2 = excitatory cell in layer 2/3, I2 = fast spiking interneuron in layer 2/3, I2L = low-threshold spiking interneuron in layer 2/3, etc. E5a and E5b are two subtypes of layer 5 pyramidal neurons.

et al., 2011c; Binzegger et al., 2004; Lefort et al., 2009) and consisted of 470 neurons in three types (excitatory pyramidal cells E, fast-spiking inhibitory interneurons I, and low-threshold spiking inhibitory interneurons IL). Each of these neural types was distributed across the four neocortical layers in the model (layers 2/3, 4, 5 and 6), giving a total of 13 distinct neuronal populations (figure 6.1). Numbers of cell in each type per layer are given in table 6.1. The population sizes and wiring densities between and within layers were taken from multiple biologically-validated sources, as documented in Neymotin et al. (2011c).

Cell type	V_{RMP} (mV)	V_θ (mV)	τ_{refrac} (ms)
I2L	-65	-47	10
I2	-63	-40	10
E2	-65	-40	50
I4L	-65	-47	10
I4	-63	-40	10
E4	-65	-40	50
I5L	-65	-47	10
I5	-63	-40	10
E5a	-65	-40	50
E5b	-65	-40	50
I6L	-65	-47	10
I6	-63	-40	10
E6	-65	-40	50

Table 6.2: Cell resting potential, threshold, and refractory period, for each cell population.

Each cell was modelled as a single-compartment integrate-and-fire neuron with fast inhibitory $GABA_A$ receptors, fast excitatory AMPA receptors, and slow excitatory NMDA receptors, each producing a voltage step in the cell membrane potential V_m followed by a decay according to a synapse-specific delay constant. For speed, the cell state variables were only updated at the time of input events, with connectivity weights obtained using a *just-in-time* strategy to avoid the requirement to hold the entirety of a massive weight matrix in memory, allowing fast and memory-efficient running even for large networks (Lytton et al., 2008). Biological behaviours such as adaptation, bursting, depolarisation blockade, and voltage-sensitive NMDA conductance were simulated using a simple rule-based system (Lytton and Stewart, 2005, 2006).

Input events at synapses depotentiated the cell membrane voltage V_m by a specified reversal potential E_{syn}, depending on the neurotransmitter type. The cell's default resting membrane potential V_{RMP} was given as a baseline, from which V_m was updated. If V_m exceeded the spiking threshold V_θ following a synaptic event, the cell emitted a spike before entering a refractory period of τ_{refrac} ms. The cell could not fire during the refractory period. This was used to set an upper limit on firing frequency. Resting potential, threshold and refractory periods for cells in each population type are given in table 6.2.

Each AMPA, NMDA, and $GABA_A$ synapse had its own voltage state V_{syn}, which

Synapse type	E_{syn} (mV)	τ_{syn}	τ_{delay} (ms)	f_{ext} (Hz)
AMPA	65	20	3–5	240–360
NMDA	90	300	3–5	40–60
$GABA_A$ (soma)	-15	10	3–5	100–150
$GABA_A$ (dendrite)	-15	20	1.8–2.2	100–150

Table 6.3: Reversal potential, V_{syn} decay constant, synaptic delay and external input frequency for each synapse type.

was added to the cell's membrane potential V_m. After each synaptic event, V_{syn} decayed exponentially with time constant τ_{syn}. $GABA_A$ events were simulated in two different ways: directly at the cell soma with a fast decay time-constant, and at the dendrite with a slower time-constant, giving two separate V_{syn} states. Synaptic events had a small delay before updating V_m, chosen uniformly from a distribution τ_{delay}.

Network activity was driven by spikes from neighbouring cells, as well as subthreshold external inputs representing activation from other areas of the brain, necessary to prevent intrinsic activity from dying out. Spikes generated by a Poisson process were provided at each synapse of each cell, maintaining average input frequencies drawn uniformly from a distribution f_{ext}. Reversal potential, V_{syn} decay constant, synaptic delay and external input frequency for each synapse type are given in table 6.3. In some simulations, an additional low-amplitude (but super-threshold) training signal was applied to layer 4 excitatory neurons (E4), representing sensory input to the neocortex.

Learning via STDP was performed at AMPA synapses between E→E cells. When a cell received or fired a spike, the interval between that event and the co-occurring spike of a neighbour was obtained, up to a limit of 40 ms maximum interspike interval. Potentiation and depotentiation then occurred by incrementing the synaptic weight between the two cells by 0.1% of baseline weight, multiplied by an exponential decay factor (time constant 10 ms) either side of the spike time, positive for potentation (*presyaptic cell fired before postsynaptic cell*), and negative for depotentiation (*postsynaptic cell fired before presynaptic cell*). It should be noted that STDP in this model uses the same increment for both potentiation and depression, whereas van Rossum argues that potentiation should instead use a multiple of current synaptic size (with depression remaining independent of current

synaptic size), making potentiation of weak synapses relatively faster than potentiation of strong synapses, as observed with experimental data (van Rossum et al., 2000). However, there is no reason to expect significant differences in qualititative behaviour for the experiments in this chapter, so the implementation was used as provided by the model.

Further details of the cell model, including activity blockade, relative refractory periods, and after-hyperpolarisation rules, in addition to wiring densities, are given by Neymotin et al. (2011c) and Neymotin et al. (2011b).

6.2.2 Synaptic scaling

The model was extended to implement synaptic scaling at E cell AMPA synapses by multiplying each cell i's postsynaptic input by a scale factor c_i, representing the multiplicative accumulation of AMPA receptors at synapses. Changes in the scale factor were calculated following the formula of van Rossum et al. (2000), with a_i as the cell's firing activity, a_i^{goal} as the target activity, β as the scaling strength, γ as the "integral controller" weight, and $\frac{dc_i(t)}{dt}$ as the rate of change of the synaptic weight:

$$\frac{dc_i(t)}{dt} = \beta c_i(t)[a_i^{goal} - a_i(t)] + \gamma c_i(t) \int_0^t dt'[a_i^{goal} - a_i(t')] \tag{6.1}$$

in which the first term modifies the synaptic weight according to the current disparity between a_i^{goal} and $a_i(t)$, and the second term allows historical under- or over-shoots of activity to pull on the weight updates more strongly, the longer they continue.

Scaling was applied inversely at $GABA_A$ synapses (i.e. by multiplying postsynaptic input by $\frac{1}{c_i}$) to enable the scaling of excitatory and inhibitory synapses in opposite directions. This allows mimicking of the effect of activity-dependent global growth factors such as BDNF, which is known to enhance activity in inhibitory circuits (Turrigiano, 2008; Chandler and Grossberg, 2012; Rutherford et al., 1998; Turrigiano, 2011). Computationally, this is subtly different from merely varying the firing threshold of the neuron. With homeostatic scaling only at excitatory synapses, an increase in E cell excitation

would lead to a corresponding increase in excitation of the inhibitory cells. This would then suppress E activity once more, and effectively counteract the scaling effect. But by scaling down the inhibitory synapses onto E cells at the same time as scaling up the excitatory synapses, the increased I cell activity (resulting from increased E cell activity) will be weighted lower by the E cells, thereby maintaining the increase in E cell activity caused by scaling, whilst also maintaining I cell firing rates.

The following parameter values were used for synaptic scaling: strength $\beta = 4.0 \times 10^{-8}$/ms/Hz; integral controller weight $\gamma = 1.0 \times 10^{-10}$/ms^2/Hz; activity sensor time constant $\tau = 100 \times 10^3$ ms. Individual cell scale factors were bounded to 100, to prevent biologically-unrealistic levels of scaling (representing AMPA receptor accumulation) up to infinity.

Average activity level for each cell i was sensed using van Rossum's slow-varying sensor $a_i(t)$, which increased with spike t_x at current timestep t, and decayed otherwise (van Rossum et al., 2000):

$$\tau \frac{da_i(t)}{dt} = -a_i(t) + \sum_x \delta(t - t_x) \tag{6.2}$$

The sensor decays exponentially as it is updated at each non-firing timestep. However, the use of event-driven just-in-time synapses in the model (Neymotin et al., 2011b; Lytton et al., 2008) meant that cell states were only updated upon each spike event rather than at every timestep, so inter-spike decay of the activity sensor could only be calculated periodically. The activity sensor was therefore modified to cope with periodic-timestep updates. Here, the first term decays the sensor according to the time between spikes $t - t_x$, and the second term increments it for the new spike, with both terms updated concurrently on the occurrence of a spike at time t_x:

$$a_i(t) = a_i(t_x) e^{-\frac{1}{\tau}(t - t_x)} + \frac{1 - a_i(t_x)}{\tau} \tag{6.3}$$

Figure 6.2 shows the activity sensor values of a simulated, randomly-spiking neuron

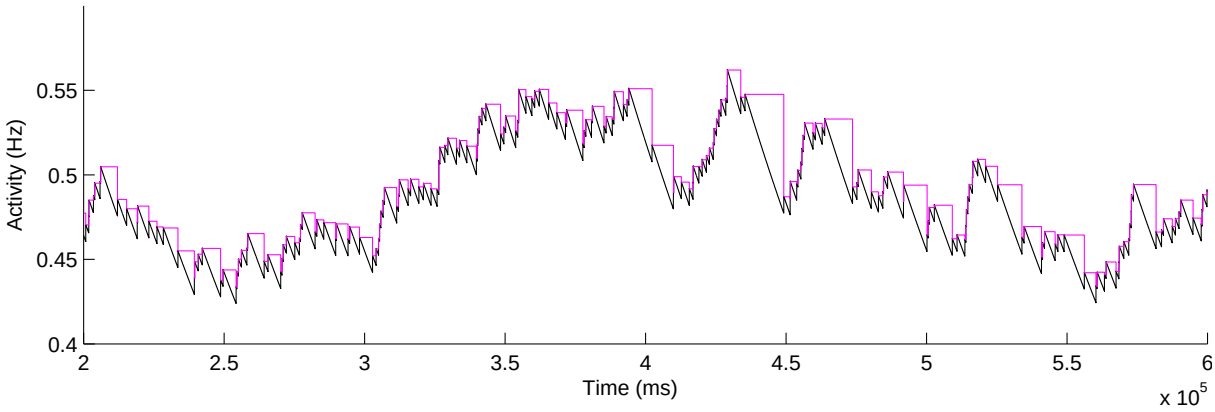


Figure 6.2: Activity sensor updating at every simulation timestep (Eqn. 6.2; black) and at every spike for activity-driven just-in-time synapses (Eqn. 6.3; magenta).

operating under the constant-timestep update policy (6.2), and the equivalent activity values under the periodic-update policy (6.3). The activity rises identically in both cases when spikes occur, but the periodic sensor does not decay until the next spike event occurs, giving the step-like appearance. The values at the spike times are therefore correct down to round-off error at the spike times.

Instead of providing an arbitrary rate target for each cell, which would fundamentally affect network dynamics, the intrinsic dynamics of the network were used to provide set-points. Initially, with synaptic scaling off, activity sensors began at 0 Hz. They were then adjusted over 800 s of simulated time based on the activity level of the cells. Synaptic scaling was then switched on.

A time constant τ of 100 s (van Rossum et al., 2000) leads to a simulation timescale of several hours for synaptic scaling: far closer to the expected biological timescale than previous studies (Turrigiano, 2008; Fröhlich et al., 2008; Chandler and Grossberg, 2012). To achieve this length of simulation, the model was extended to allow periodic flushing of all spike data to disk, enabling very long runs (unlimited except for available disk space). A typical simulation of 44 hours ran in approximately real time and produced around 2 GB of spike data. The model was implemented in NEURON 7.2 (Carnevale and Hines, 2006) for Linux, and is available on ModelDB at the following URL: `https://senselab.med.yale.edu/modeldb/enterCode.asp?model=147141`.

6.3 Results

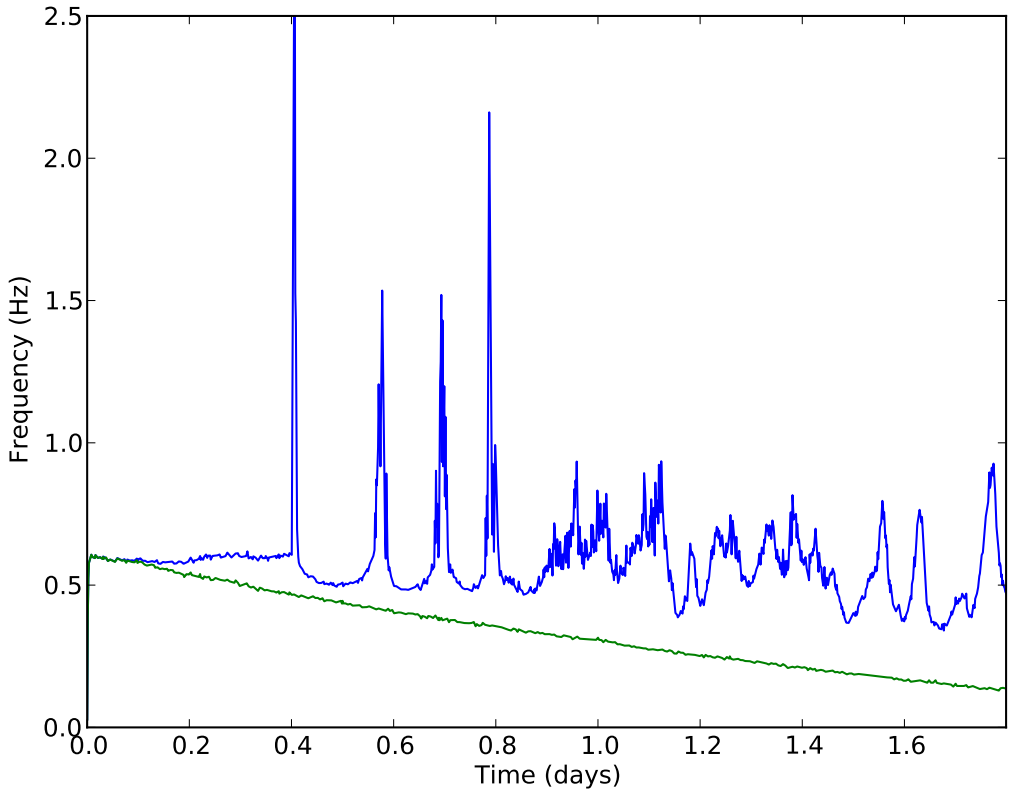
The following experiments were tested over 20 runs with different random seeds on each run for initial synaptic weights, cell positioning, cell wiring, and external input spike trains. Representative sample figures from an individual run with all random seeds set to 1 are shown alongside mean figures for all 20 runs.

6.3.1 Scaling prolongs activity during deletion

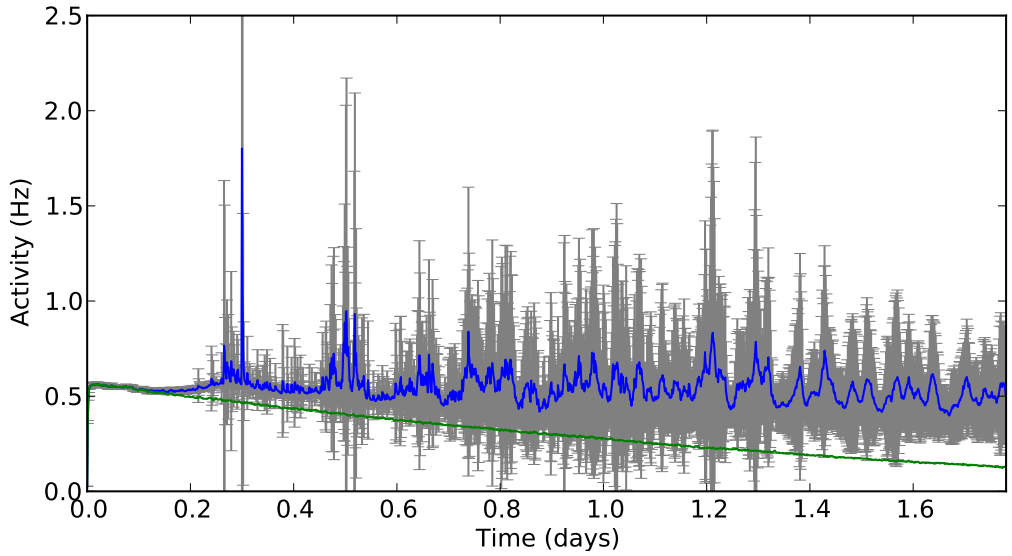
As discussed, synaptic scaling may be useful for compensating for decreases in activity during atrophy, for example in degenerative diseases. To show that synaptic scaling can maintain activity, the network dynamics were altered through gradual removal (pruning) of cells (figure 6.3) over a period of 160,000 s ($\approx$ 44 h). Every 1600 s during the simulation, three I or E neurons were selected at flat-random and removed from the network by setting all of their synaptic weights to zero. The global external input weights were scaled down proportionally to the total amount of deletion, at a quarter of the deletion rate, to prevent the external inputs from swamping internal activation and artificially raising activity. By the end of the two-day simulation, approximately two thirds of the cells in the network had been deleted.

In the absence of scaling, average firing rates across E cells declined steadily as cell deletion progressed (figure 6.3 green / lower line). With scaling present, firing activity was maintained (figure 6.3 blue / upper line). Initially the activity is entirely stable, but after this there are brief activity peaks caused when the inherent delay in the activity sensor led to over-compensation. These activity peaks, separated by several hours, do not correspond to deletion times (which occur every 1600 s), but rather to emergent instabilities in the resulting damaged network. Indeed, the network remains stable for nearly half a day following the onset of deletion after 800 s.

The over-compensation can be adjusted to some degree, although not eliminated, by altering the scaling parameters β and γ (not shown), but the activity peaks appear to



(a) Sample run: E activity without scaling (green/lower), with scaling (blue/upper)



(b) Mean activity over 20 runs

Figure 6.3: E activity during pruning with (blue / upper) and without (green / lower) compensatory synaptic scaling. Run time 160,000 s ($\approx$ 44 h).

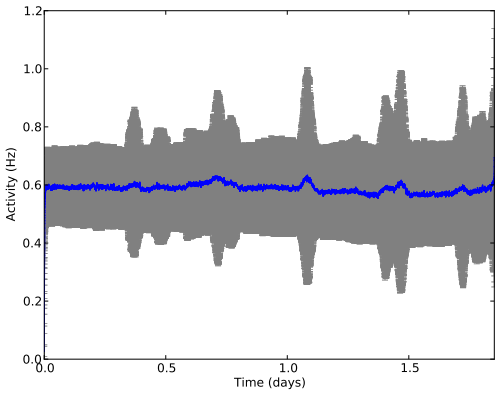
be a fundamental property of scaling in spiking neural networks (Fröhlich et al., 2008). There is evidence that such activity peaks may have a biological basis, representing the increased incidence of epileptiform bursting in individual neurons and in whole networks in Alzheimer's disease (Busche et al., 2008; Fröhlich et al., 2008; Trasande and Ramirez, 2007), and this could be a mechanism for excitotoxic cell death in Small's hypothesis of Alzheimer's disease progression (Small, 2008).

6.3.2 Synaptic scaling does not disrupt network behaviour

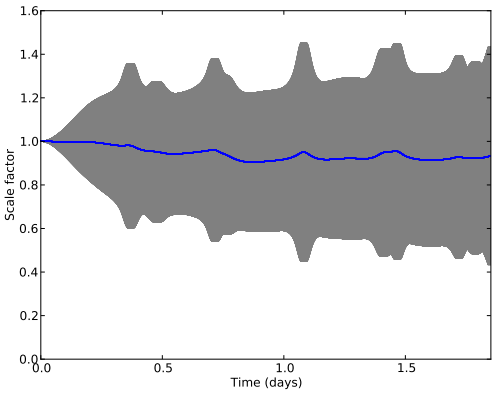
To investigate whether this implementation of synaptic scaling is inherently unstable and directly responsible for these activity peaks in a normal network, or whether the activity peaks are simply an emergent network response to compensatory synaptic scaling, the model was run for 160,000 s ($\approx$ 44 h) to examine the effects of scaling over time on network dynamics in the absence of any other external effects. With scaling, activity of the E cells remained steady without any activity peaks (figure 6.4a), and average scale factors remained centered around 1 (figure 6.4b). Therefore it appears that synaptic scaling is sufficiently stable to maintain normal activity in a network of spiking neurons during extremely long simulation runs, and is not directly responsible for the instability seen during the previous experiment, but rather that the instability is only seen during periods of damage.

6.3.3 Unrestrained STDP leads to hyper-potentiation

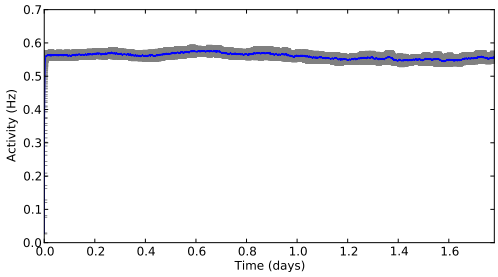
To investigate the need for scaling during learning, the network was trained according to the schema of Neymotin et al. (2011b) by applying a signal consisting of low-weight single spikes at 8 Hz to E4 cells for 8000 s ($\approx$ 2.2 h) in the absence of synaptic scaling. STDP was turned off during the final 800 s in order to show the network dynamics during a stable "recall" policy. It can be seen figure 6.5 that the training signal eventually pushed the network into a state of excessive firing via runaway hyperpotentiation of synapses, with



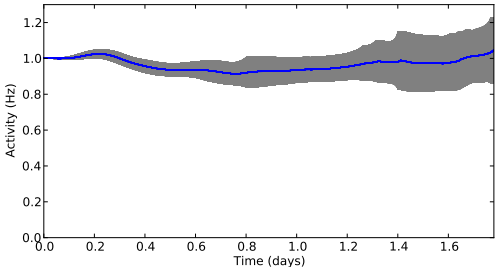
(a) Sample run: E activity during scaling



(b) Sample run: scale factors of E cells



(c) Mean activity over 20 runs



(d) Mean scale over 20 runs

Figure 6.4: Scaling does not destabilize the network (mean: blue, std: grey).

a sample raster image in figure 6.5c displaying repeated high-frequency waves of activity between cells. This occurred even when the manual E→I STDP balancing described in Neymotin et al. (2011b) was added (not shown).

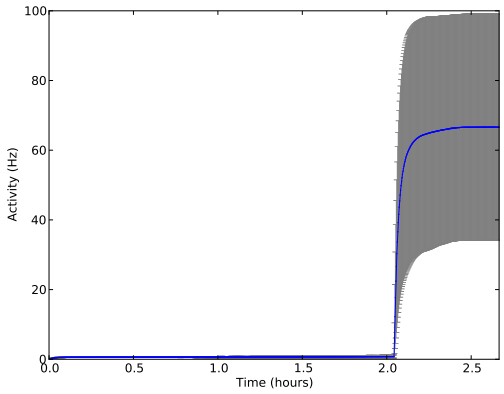
6.3.4 Synaptic scaling prevents overactivation

To determine whether synaptic scaling is capable of balancing this hyperpotentiation during learning, the model was run with STDP and synaptic scaling activated, again with an 8 Hz training signal. Figure 6.6 shows stable activity during the learning process, with a gradual decrease in the average scaling factor over time as the cells compensated for the increased potentiation from learning. Local E cell homeostatic scaling balanced the potentiation caused by STDP, gradually scaling down all E cells, and preventing pathological over-activation.

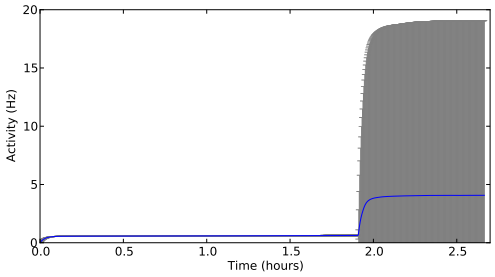
6.3.5 Synaptic scaling preserves learning

It has been demonstrated that synaptic scaling can maintain cell firing near the target rate, here set at the baseline firing rate for each cell. However, successful learning depends on the relative alteration of innervating synapses on each cell. It is possible that the scaling-down of activity simply globally (in terms of each neuron) reverses the potentiation caused by STDP, resulting in a loss of learned information. In order to determine whether scaling allowed the learning of oscillations to persist, power spectra of the E cells were obtained during the learning process via Fourier analysis of the population spike trains.

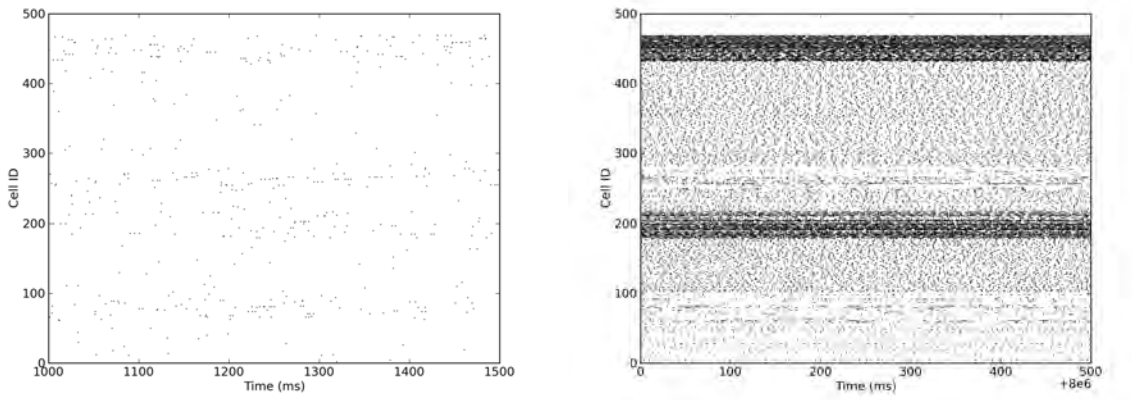
Simulation spike trains were organized into multi-unit activity (MUA) vectors, defined for a cell population as the number of spikes in the population over a time interval (bin). Bin sizes were set to 5 ms (giving a 200 Hz sampling rate). Analysis was performed by subtracting the mean from each of the MUA vectors, with power spectra calculated using the multitaper spectral power estimation method, as implemented in the Python wrapper of the FORTRAN MTSpec library (Prieto et al., 2009) (see figure 6.7a for an example).



(a) Sample run: E activity during training

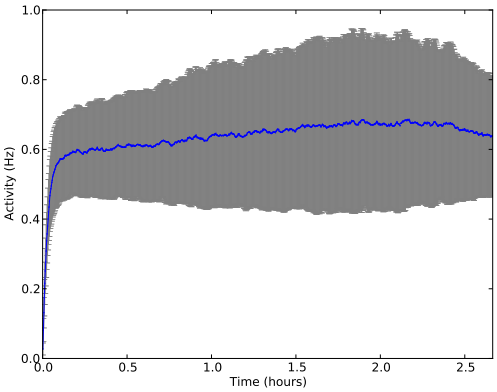


(b) Mean activity over 20 runs

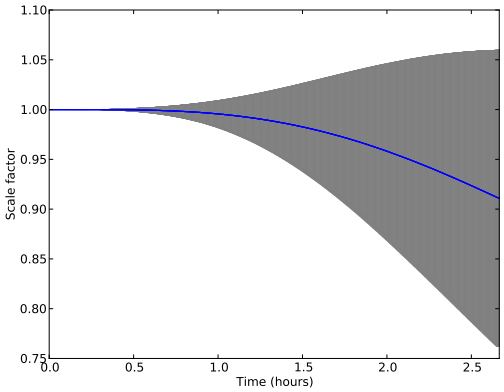


(c) Sample raster plots of 500 ms showing normal activity before (left) and high-frequency whole-network activity after (right) 2 h of training

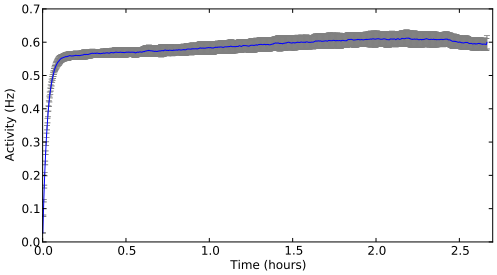
Figure 6.5: Training with E→E STDP pushes the network to high frequency activity.



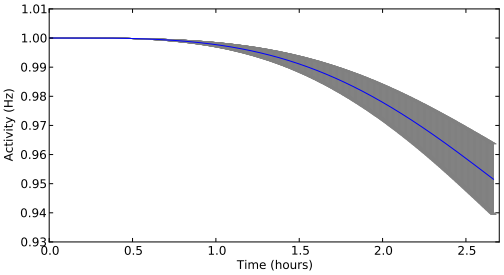
(a) Sample run: E activity during scaled training



(b) Sample run: scale factors of E cells



(c) Mean activity over 20 runs



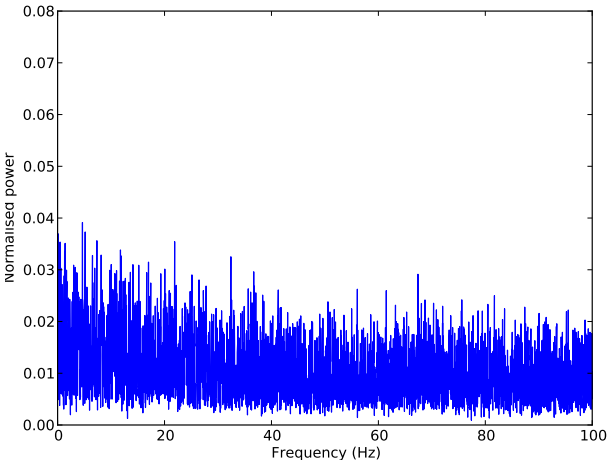
(d) Mean scale over 20 runs

Figure 6.6: Synaptic scaling maintains E activity profile during STDP.

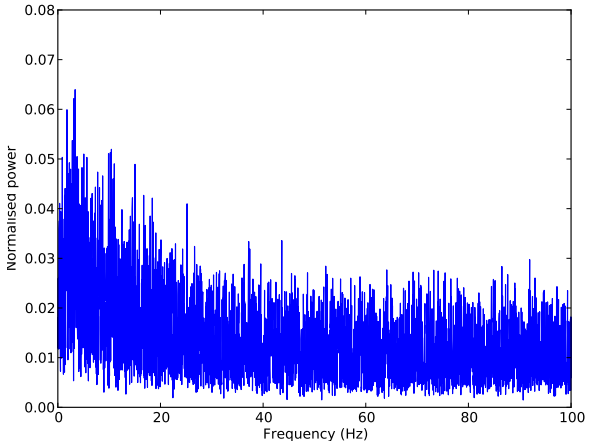
These plots show unsmoothed normalised power of the E cells within the network at each of a range of frequencies from 0-100 Hz, indicating the frequencies at which the network oscillations are strongest.

STDP was applied at E→E synapses for 8000 s (≈ 2.2 h) with an 8 Hz sensory signal (figure 6.7). In one simulation, synaptic scaling was also switched on for E cells. Power spectra were obtained for the period from 5600-6400 s, shortly after the middle of training (figures 6.7b and 6.7c), and again during the recall period at the end of learning in which STDP was switched off (figures 6.7d and 6.7e).

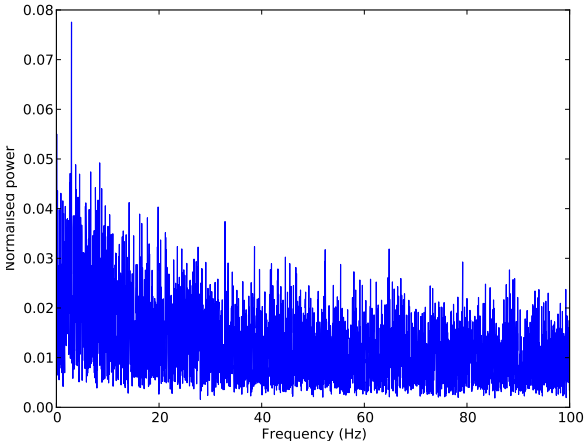
In both simulations, it can be seen that STDP has caused a shift in the power spectra, with an increase in the amplitude of oscillations at low frequencies from 0-10 Hz (figures 6.7b and 6.7c). This demonstrates that the network has learned from the training signal. Shortly after 7400 s (2 h), the network without synaptic scaling transitioned to high-frequency activity, without retention of the 8 Hz training signal (figure 6.7d; note different



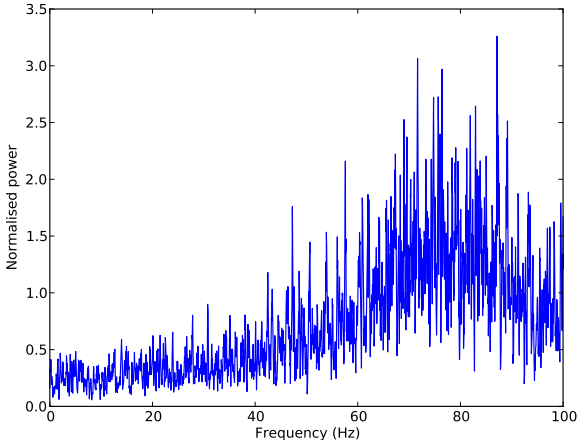
(a) Baseline power spectrum of E cells (frequency (Hz) vs normalized power)



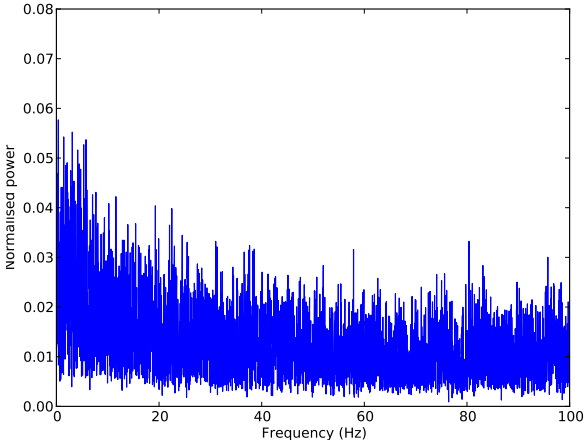
(b) With STDP only, during training



(c) With STDP and scaling, during training



(d) With STDP only, after training



(e) With STDP and scaling, after training

Figure 6.7: Power spectra during (top) and after (bottom) a 2-hour period of learning via STDP, with (right) and without (left) scaling.

scale). However, in the network with synaptic scaling turned on, lower frequency activity was maintained, with a peak near the 8 Hz that was imposed during training (figure 6.7e). Synaptic scaling therefore prevented over-activation and preserved learning.

6.4 Discussion

This chapter has introduced homeostatic synaptic scaling with dynamically-obtained target activity rates to a realistic spiking model of neocortex which learns oscillatory frequencies via STDP. It has shown that scaling is both necessary and capable for upregulation of neural activity during decline in input, for example during the progression of neurodegenerative brain disorders such as Alzheimer's disease, in which cortical activation might be expected to decrease. Peaks of synchronised activity were observed during deletion due to periodic over-compensation by the scaling mechanism. Experimental observations demonstrating hyperactivity in cells near beta-amyloid plaques in Alzheimer's disease, and the increased incidence of seizures in Alzheimer's patients, suggests these activity peaks may have a biological basis (Busche et al., 2008; Fröhlich et al., 2008; Trasande and Ramirez, 2007).

It was also demonstrated that scaling does not negatively affect the network at baseline, but that it is inherently stable, and that scaling between E cells is sufficient to balance the hyper-potentiation caused by unconstrained STDP. Potentiation strengthens the co-incident connections between neurons in a positive feedback cycle, eventually leading to hyper-potentiation, but scaling acts to shift the mean activation continuously back towards the target activity. At the same time, the relative (learned) distribution between postsynaptic weights remains unaltered by scaling, and this principle was subsequently demonstrated by showing that learning of an 8 Hz oscillatory signal is not erased by scaling.

This model investigated learning and scaling at E→E synapses between E cells (with additional opposite scaling at inhibitory synapses onto E cells). Whilst there is also some

evidence of STDP between I→I inhibitory cells (Lamsa et al., 2010), inhibitory cells do not appear to perform scaling, but rather: "homeostatic regulation of inhibition is a noncell-autonomous process that either requires changes in both pre-and postsynaptic activity simultaneously or is triggered by global changes in network activity" (Turrigiano, 2011). In this model, directly enabling synaptic scaling in I cells was found to lead to dramatic instabilities in the network dynamics (even when operating the network at baseline, i.e. without STDP or a sensory signal), which is consistent with Turrigiano's observations. Rather, the network appears to be most stable when I cells are allowed to adjust their activity passively according to the changing output from neighboring E cells, thus requiring only one dimension for the E/I balance rather than needing a second simultaneously active dimension for scaling.

STDP was implemented using an incremental step of 0.1% of baseline synaptic weight, which may seem low. Increasing this step size, however, meant that short bursts of high-frequency activity were seen during learning, as the activity sensors could not respond quickly enough to cause sufficient compensatory scaling (although the network did soon scale back to previous firing rates). However, 8000 s (2 h) of sustained training may also be very long compared to biological learning from hippocampal backprojections, which is known to include periods of recall and consolidation between periods of learning (McClelland et al., 1995).

6.5 Chapter summary

This chapter has introduced homeostatic synaptic scaling to a biologically-realistic spiking neural network simulation, and extended the simulation to allow for extremely long runs (effectively unlimited, but here typically two days of simulated time). This is longer than has been achieved in similar previous models. The activity sensor of van Rossum et al. (2000) was also modified to work with periodic updates instead of a fixed timestep, as required by modern simulators' fast just-in-time synapses.

It has been demonstrated that the scaling rule is stable in the absence of network damage, and that it can maintain activity during network damage caused by random neuronal deletion, but that it subsequently results in slow-periodic peaks of over-activity. This is a new contribution to knowledge, both describing a possible toxicity mechanism for Small's scaling-driven progression hypothesis of Alzheimer's disease (Small, 2008), and offering an explanation for the medically-observed hyperactivity of neurons near amyloid plaques in Alzheimer's disease.

This chapter has shown, using a biologically-realistic cortical columnular model operating on realistic timescales for the first time, that synaptic scaling is capable of balancing hyperpotentiation during learning via STDP, but that it does not erase learned shifts in oscillatory dynamics during this process. It has also demonstrated for the first time that the stability of the synaptic scaling mechanism is disrupted during Alzheimer's-like atrophy, giving rise to epileptiform bursting, which matches medical observations.

CHAPTER 7

INFORMATION-SELECTIVITY IN A SPIKING MODEL

Chapter synopsis

Chapter 6 showed that synaptic scaling in a biologically-realistic spiking model of neocortex plays an important role during learning, balancing hyperpotentiation and preventing runaway epileptiform activity. However, Small (2008) predicts that synaptic scaling may also drive the progression of Alzheimer's disease pathology, targeting neurons which scale up to compensate for reductions in activation.

Previously, in chapter 4, a hypothesis was proposed which predicted that the neurons with the weakest connectivity, and therefore the lowest contribution of information to the network, would be those which scale up the most and are the first to die in early stages of Alzheimer's disease. This would cause a period of cell death from within the 'cognitive reserve', with a corresponding lack of cognitive symptoms, thus masking the presence of the disease and making timely intervention and treatment more difficult. In this chapter, the predictions of this hypothesis are tested in a more biologically-realistic model of neocortex.

The Yu et al. (2010) method (summarised in Crumiller et al. (2011)) is used to obtain information contribution measures from the network, and a regime of stimulation which successfully probes the synaptic weights is developed. The results of this method are used

to show that the network has distinct levels of information contribution in each of its different layers, matching predictions by Neymotin et al. (2011c).

Finally, the information-selectivity hypothesis is tested by lesioning the network with a representative range of different cell death strategies: cells are deleted either at uniform-random, or in proportion to the deviation of their scaling factors from their starting values, or in proportion to the amount and length of time of activity above the target firing rate. Synaptic scaling is additionally accompanied in some simulations by a global neurotrophin-driven scaling signal, modelling the effects of BDNF and TNF-α, further enhancing the effects.

7.1 Introduction

Small's hypothesis for the progression of Alzheimer's disease (AD) as driven by compensatory synaptic scaling (Small, 2008) leads to predictions that the greatest scaling will occur in sparsely-connected neurons, as a function of their greater relative decline in activation per deleted postsynaptic innervation (Rowan, 2012). A further prediction is that these sparsely-connected neurons will, as a result of their reduced connectivity, contribute less information to the network. Therefore, low-significance neurons will scale up more in early stages of AD than high-significance neurons, meaning that the symptomatic effect of cognitive decline will be delayed until later after disease onset, thus masking the presence of the disease and making timely intervention more difficult.

These predictions were tested in chapter 4 in a simple associative Hopfield-class model with many abstractions, omitting such details as networks of spiking neurons, and their arrangement in a layered architecture such as is found in the neocortex. In order to test this hypothesis more reliably, it is necessary to use a more biologically-realistic model which addresses some of the simplifications made in previous experiments.

An existing spiking neural model of a neocortical column which displays biologically-validated emergent oscillations, and is wired according to biological data (Neymotin et al.,

2011c), is extended to incorporate local homeostatic synaptic scaling according to the schema of van Rossum et al. (2000), and activity-dependent cell death. The dynamics of synaptic scaling are additionally enhanced to simulate the effects of the neurotrophin BDNF and the global growth factor TNF-α, allowing a combination of local and global scaling effects to occur.

Obtaining estimations of information contribution in spiking neural networks is not an easy task (Quiroga and Panzeri, 2009), but a method using Fourier analysis can be applied to the output of the network under repeated trials of repeated and unique stimuli to obtain per-cell, sum-total, and combined group information measures (Crumiller et al., 2011). The network must be stimulated in a specific way to elicit the required responses, otherwise the information calculation will be dominated by noise and will provide incorrect results. Therefore, initial experimentation is performed to ascertain a usable regime of stimulation for the network.

Following this, the progression hypothesis of AD is tested in the model by lesioning the network with uniform-random, scaling factor-dependent, and activity-dependent cell death in the presence of compensatory synaptic scaling, with and without additional global neurotrophic scaling signals. The resulting patterns of cell death are compared with the information-contribution measures obtained via the Crumiller et al. (2011) method, in order to test the validity of the hypothesis of information-selectivity of AD pathology.

7.2 Methods

7.2.1 Model description

The spiking neural model used the same architecture, wiring parameters, and synaptic scaling mechanism as given in the previous chapter (section 6.2), but with additional features to explore the hypothesis of information-selectivity within the network. A description of these additional features is given in the following sections.

7.2.2 Synaptic scaling driven by neurotrophic factors

Small (2008) predicts that not only should cells homeostatically scale back up to their target activity in order to compensate for lost activation, but that the network as a whole should also scale up its activity in response to damage, in order to maintain a global absolute firing rate and therefore maintain information processing throughputs. Individual cells are therefore required to alter their activity targets, so they can scale up beyond their original 'design' firing rate to compensate for the global loss of activity. Such a mechanism would require scaling to be goverened by a global activity signal in addition to the local signal generated by each cell's activity sensor. This global signal takes the form of a combination of a reduction in the release of the activity-dependent neurotrophin BDNF, and an increase in the release of the growth factor TNF-α from neighbouring glial (support) cells (Turrigiano, 2008).

Neurotrophic scaling was incorporated into certain simulations by calculating a global scaling signal C as a proportion of global activity A, and comparing this to the baseline target global activity A^{goal}:

$$A(t) = \sum_i a_i \tag{7.1}$$

$$C(t) = \frac{A^{goal}}{A(t)} \tag{7.2}$$

Synaptic scaling for each cell (equation 6.1) was then adjusted to include the global scaling signal C as a multiplier of the 'design' target activity a_i^{goal}:

$$\frac{dc_i(t)}{dt} = \beta c_i(t)[C(t) \cdot a_i^{goal} - a_i(t)] + \gamma c_i(t) \int_0^t dt'[C(t) \cdot a_i^{goal} - a_i(t')] \tag{7.3}$$

This had the effect of allowing the target activity level of a cell to be adjusted dynamically up and down in response to global scaling signals, as hypothesised by Turrigiano (2008), and discussed in section 5.3, but with the scaling mechanism still applied locally at each cell based on changes in its postsynaptic input.

7.2.3 Cell death

Random baseline

To provide a baseline with which to compare further experiments, deletion was performed by selecting a fixed number of cells at random at each deletion timestep. In the following experiments, 3 inhibitory or excitatory cells were picked at random every 1600 s, over 100 rounds of deletion during the simulation, giving a total of 300 deleted cells (around 64% of the network). Synaptic scaling was then used to bring each excitatory cell back to its target firing rate.

Scaling factor-proportional

Small (2008) predicts that cell death is a function of over-activity related to scaling factors. For this reason, as in the earlier associative Hopfield-class model (chapter 3), cell death was modelled in proportion to scaling factor values, as an abstraction of an unspecified biological method of excitotoxicity.

In a network which compensates for hypoactivity by increasing the scaling factors of individual neurons, it may be reasonable to assume that, across the network, *all* cells will scale up – and this is true on average. However, in the spiking model, it was observed that some cells were constantly driven to over-activity by other cells to which they were connected (shown later in the results), and were actually forced to scale *down* in order to maintain their target firing rates. These cells are also likely to be susceptible to excitotoxicity, as they are constantly being driven to activity above their target firing rates.

Due to the observation of this forced over-activation (and hence, the greater susceptibility to excitotoxicity) of some cells during scaling, cell death was modified to take these down-scaling cells into account. The probability of deletion was therefore made proportional to *deviations in scaling* from the starting value of 1, rather than it simply being proportional to the absolute scaling value.

This means that, for two example cells, if one cell is forced towards over-activity and excitotoxicity, and subsequently scales down to $c = 0.5$, its deviation from $c = 1$ is a factor of two. This cell would therefore be deleted before another cell which may have scaled up by a third to $c = 1.33$ due to hypoactivity, as the implication is that the first cell suffers more excitotoxic damage than the second. Conversely, two different cells with scale factor values of $c = 0.1$ and $c = 10$ both have the same level of deviation from 1, and would therefore have the same probability of deletion as each other.

At each deletion timestep (again, every 1600 s), the 3 cells with the greatest deviations of scale factor from $c = 1$ were therefore chosen and removed from the network.

Excitotoxicity

Deletion based on some abstract notion of scaling factor is not necessarily biologically realistic. Small (2008) hypothesises that, as cells are driven to over-activity beyond their target firing rates, the subsequent influx of activity-dependent calcium ions – exacerbated by the β-amyloid pathology present in AD (Demuro et al., 2010) – triggers an apoptopic cell-death mechanism. The cell is programmed to die safely and cleanly in order to prevent more damaging *necrosis*, in which the cell dies, its membrane breaks down, and parts of the dead cell are leaked into the intracellular space, provoking further damage.

Such over-activity may be caused by excessive activation following compensatory scaling beyond a safe limit, or by activity peaks caused by synchronised network-wide bursting (as seen in the previous chapter in figure 6.3a), or by cells being uncontrollably driven by strongly-innervating neurons to which they are connected.

In simulations with excitotoxicity, cell death for inhibitory and excitatory cells was modelled by calculating, at the time of each synaptic event, a probability of death P_{death}. This value was determined by the current activity level of the cell with respect to its target activity level, multiplied by a deletion rate constant τ_{del}, and multiplied by the

time since the last synaptic update (equation 7.4):

$$P_{death}(t') = \tau_{del} \frac{a_i(t') - a_i^{goal}}{a_i^{goal}} (t' - t) \tag{7.4}$$

This mechanism of deletion also allowed variable rates of deletion driven directly by the pathology, rather than imposing a fixed timestep between rounds of deletion, making it possble to experimentally observe changes in the rate of deletion caused by various effects of the pathology, in addition to observing the information-related effects.

In order to start the feed-forward process of cell death, it was necessary to provoke compensatory up-scaling, mimicking the onset of Alzheimer's disease. This was achieved by reducing the weighting of the external inputs to each cell globally by 5% of their original values, causing cells to scale up above baseline levels, thereby enabling further Aβ-mediated cell death and the onset of the damage cascade.

In the following experiments, τ_{del} took the value 1×10^{-4}, giving a level of deletion of approximately $50 \pm 20\%$ of cells over two days of simulated time. Altering τ_{del} was found to have only linear effects on the global deletion rate. For experiments with an additional global neurotrophic scaling signals, τ_{del} was reduced to 1×10^{-5} to account for the forced increase in activity of each cell and subsequent higher probability of cell death. This allowed the simulation to continue to achieve a similar level of cell death over two days, allowing easier comparison of results.

7.2.4 Associative pattern learning via STDP

In some experiments it was necessary to enable associative learning of certain discrete firing patterns. Patterns were stored as unique vectors of approximately 15% of E cells which would be stimulated to elicit the relevant response from the network. 31 patterns (30 for the 'unique' stimuli and one for the 'repeat' stimulus) were created, with each E cell having a 15% chance of being selected as part of each pattern.

During pattern storage, scaling targets were obtained for each cell from 1600 s of

baseline activity, before STDP was switched on for 3100 s (≈ 1 h). During this learning period, patterns were selected at random and presented to the network via the external inputs at above-threshold strength for 8 s, with a further 8 s without stimulation between presentations to allow the network to recover to baseline. This gave an average of 6.25 presentations of each pattern during the learning process.

In some further experiments involving pattern learning, cholinergic "clamping" (section 2.1.2) was applied to cells which were not part of each pattern, in order to prevent spurious reactivation of previously-learned patterns during learning (Hasselmo, 1993). This was implemented in a simplified manner by stimulating the inhibitory $GABA_A$ synapses of the relevant non-pattern cells, in order to prevent them from activating during the pattern presentation and to facilitate learning of new basins of attraction.

7.2.5 Information contribution

Fourier information measure

As described in section 5.4, information contributions for each neuron, and for the network as a whole, were calculated using a Fourier analysis approach (Yu et al., 2010; Crumiller et al., 2011). The network was tested over 100 trials, each consisting of presentation of a 'repeat' stimulus for 8 s, followed by one 'unique' stimulus, which was either generated randomly or drawn from a set of pre-defined stimulation patterns, for 8 s.

Spike times were then used to define Fourier coefficients in the frequency domain for each trial, using the algorithm of Yu et al. (2010). At each frequency, there are therefore 100 Fourier coefficients (one for each trial). The entropy contained within the responses is calculated by taking the log of the variance of this distribution. The entropy relating to the 'repeat' stimulus in each trial was subtracted from the entropy for the 'unique' stimulus in the trial, with the resulting difference taken as the information rate in bits/s for each frequency.

If correctly set up, the presentation of alternating 'repeat' and 'unique' stimuli for

8 s each should result in graphs of information rate (bits/s) vs frequency (Hz), in which information contribution from each cell monotonically increases at decelerating speed from 0 Hz to a maximum between 100-200 Hz. This levelling-off is predicted by Yu et al. (2010), due to the entropy of the 'repeats' and 'uniques' trials tending towards the same value, and therefore the difference between them tending towards zero.

In cases where the noise entropy of the network cannot be reliably computed, for example when a cell does not respond in a similar way to each of the presentations of the 'repeat' stimulus, the information contribution of that cell will remain around zero, or in some cases even be negative. This is clearly an anomalous result (there is no such thing as negative information), but it indicates that the neuron's reliability of response to the stimuli is very low.

In the case of the individual neuron, this may be because the neuron is just responding randomly to all the stimuli, in which case its information contribution is indeed close or equal to zero. But if the whole network should show zero or negative information across the frequency spectrum, then it indicates a problem with the stimuli themselves which is preventing the network from responding in a reliable manner.

The information calculation measure assumes that the 'repeat' stimulus is representative of the set of 'unique' stimuli, and that the resulting spike trains therefore have similar statistical properties to each other. From this, it can be deduced that the 'repeat' and 'unique' stimuli must be carefully defined in order to elicit a reliable response. Section 7.3.1 shows the results of experiments using different types of stimulation to attempt to obtain reliable information contribution measures.

Verifying network response reliability using Kernel Density Estimation

The information calculation algorithm requires a reliably similar response to the stimuli over all 100 trials, so that noise can be calculated and subtracted from the overall result. One way to determine the degree to which each neuron in the network responds similarly to the stimuli is to use the Kernel Density Estimation (KDE) measure.

KDE allows a non-parametric estimation of the probability density function of a random variable. Here, the random variable is defined by the spike times of each neuron in response to the 'repeat' and 'unique' stimuli, over all 100 trials. The resulting estimated probability density function can be thought of as a normalised, continuous histogram over the spike times. Figure 7.1 shows an example Kernel Density Estimation of the responses of two neurons driven by an unknown process, in which one of the neurons responds much more reliably than the other.

The reliability of response can additionally be quantified using the Coefficient of Variation (CV) measure, which is simply a unitless normalised measure of the dispersion of the probability density, obtained as the ratio of the standard deviation to the mean:

$$c_v = \frac{\sigma}{\mu} \tag{7.5}$$

In the following experiments, KDEs were obtained using freely-available *kernel density estimation by diffusion* MATLAB code (Botev et al., 2010).

7.3 Results

7.3.1 Obtaining positive information measures

In order to obtain valid information contribution measures for each neuron in the network using the Crumiller et al. (2011) method, a number of different regimes of 'unique' and 'repeat' stimulation were tried. The information calculation was run using the resulting spike trains for each excitatory cell in the network over 100 trials of alternating 'unique' and 'repeat' stimuli. Each experiment was repeated over 20 runs to provide mean and standard deviations of average E cell firing rates, average CV measures for each stimulus set, and average information contribution of E cells in bits/s. Raster plots were generated, showing the time of spikes from each cell (marked as a dot), to display general effects such as network-wide bursting or changes in mean firing rates as stimulation varied.

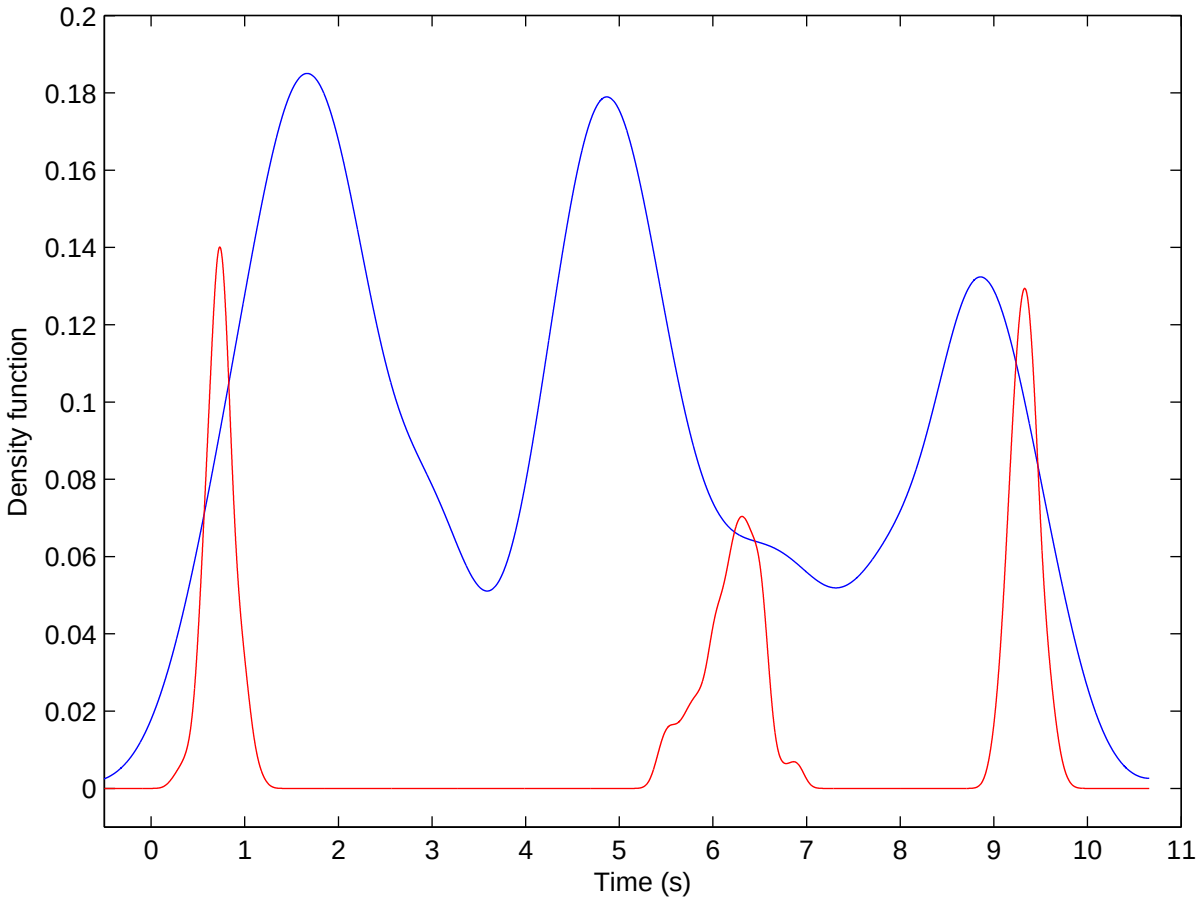


Figure 7.1: Kernel Density Estimations for two neurons. One (blue/upper) responds only semi-reliably over multiple trials ($c_v = 0.5747$), with other spikes distributed randomly across the time domain. The other neuron (red/lower) responds much more reliably ($c_v = 2.1402$), as evidenced by the near-zero probability density between spike times.

Varying frequency of stimulation

The model was driven for the 'repeats' stimulus with above-threshold stimulation at 16 Hz applied to 80% of E4 cells (excitatory pyramidal cells in layer 4; the sensory 'input' layer for the neocortex) which were picked at random but fixed across trials. The 'uniques' stimuli were provided by varying the phase and frequency of this stimulation between 4 Hz and 20 Hz, and varying the 80% of cells picked on each trial.

It was found that stimulation of 80% of E4 cells at any one time was sufficient to generate activity across the whole network, but stimulating larger areas of the network simultaneously had a tendency to drive the network to high-frequency bursting (not shown). Stimulation even below 20 Hz also tended to push the network to bursting (fig 7.2).

The large differences in mean E cell firing rates between the repeat trials and the unique trials (table 7.1) are a result of the variation in stimulation frequencies, and are liable to increasing noise in the information calculation. The CV measure for the repeat trials is reasonably high, indicating that the network was responding in a self-similar way to the 'repeat' stimuli across the 100 trials, but the CV measure for the unique trials is much lower, which again increases the disparity between repeat and unique trials, and therefore increases noise in the calculation.

This resulted in an information calculation which was subject to excessive amounts of noise, as can be seen in the resulting group information graphs (fig 7.3), and from the fact that the mean information of all cells in the group was not very high, with many runs showing less than 1 bit/s on average.

Stimulation of random patterns

To attempt to equalise the firing rates between 'unique' and 'repeat' stimulation, it is clearly necessary to use a similar frequency of stimulation for each of the trials. Therefore, if frequency of stimulation cannot be used to distinguish each unique trial from the others, it may make sense to stimulate differing subsets of cells using a fixed frequency.

In order to allow sufficiently diverse sets of discrete patterns, and in recognition of the

Firing rate (Hz)		KDE Coefficient of Variation		Info (bits)
Repeats	**Uniques**	**Repeats**	**Uniques**	
3.21 ± 3.16	2.71 ± 2.72	0.8848 ± 0.3228	0.8255 ± 0.3537	1.34 ± 1.82
3.30 ± 4.31	1.43 ± 0.87	0.9465 ± 0.3006	0.5556 ± 0.1324	6.33 ± 5.54
3.91 ± 3.15	3.21 ± 1.89	0.9892 ± 0.4823	0.4621 ± 0.0631	11.84 ± 12.06
9.59 ± 9.46	3.35 ± 8.38	1.1767 ± 0.3466	0.6116 ± 0.2362	0.08 ± 0.43
13.07 ± 16.25	3.22 ± 5.05	1.0693 ± 0.3550	1.6763 ± 0.8878	0.70 ± 1.88
2.49 ± 2.16	1.53 ± 0.70	1.1487 ± 0.4372	0.4954 ± 0.1126	12.96 ± 12.92
3.90 ± 8.63	2.29 ± 8.19	0.9053 ± 0.2302	0.4040 ± 0.0302	0.01 ± 0.14
3.63 ± 4.18	1.70 ± 1.16	1.3429 ± 0.6361	0.6251 ± 0.3432	4.37 ± 5.07
3.41 ± 3.41	1.60 ± 0.83	1.4133 ± 0.6513	0.6551 ± 0.2317	7.18 ± 7.14
8.75 ± 5.46	1.62 ± 0.66	1.3480 ± 0.4433	0.5752 ± 0.1873	2.57 ± 2.16
2.50 ± 2.57	1.58 ± 0.73	1.0490 ± 0.4418	0.4828 ± 0.1080	13.18 ± 12.70
3.82 ± 3.86	1.73 ± 1.00	1.0681 ± 0.3800	0.8138 ± 0.3856	1.67 ± 2.82
3.84 ± 3.45	1.64 ± 0.70	1.5748 ± 0.7408	0.6818 ± 0.2695	9.86 ± 9.28
4.47 ± 5.38	2.13 ± 5.47	1.2446 ± 0.4708	0.5540 ± 0.1347	0.05 ± 0.28
2.71 ± 1.89	1.87 ± 1.38	1.3604 ± 0.6819	0.7843 ± 0.3195	0.96 ± 2.19
7.92 ± 8.04	2.67 ± 5.17	1.1295 ± 0.3480	0.7642 ± 0.2434	0.12 ± 0.47
3.78 ± 3.87	1.61 ± 0.89	1.0597 ± 0.3919	0.9450 ± 0.4706	2.88 ± 4.59
4.22 ± 6.28	2.58 ± 5.48	1.1001 ± 0.3788	0.7818 ± 0.2727	0.05 ± 0.28
2.38 ± 1.29	1.63 ± 0.61	1.4026 ± 0.5889	0.4513 ± 0.1166	12.79 ± 12.31
10.93 ± 6.68	2.67 ± 2.95	0.9164 ± 0.2042	0.7180 ± 0.2298	0.35 ± 1.01
5.09 ± 3.13	2.14 ± 0.64	1.1565 ± 0.1960	0.6931 ± 0.2741	4.46 ± 5.00

Table 7.1: Mean firing rate of E cells over 100 'repeat' and 'unique' trials during varying frequencies of stimulation, with mean KDE coefficient of variation and mean total information, for each experimental run.

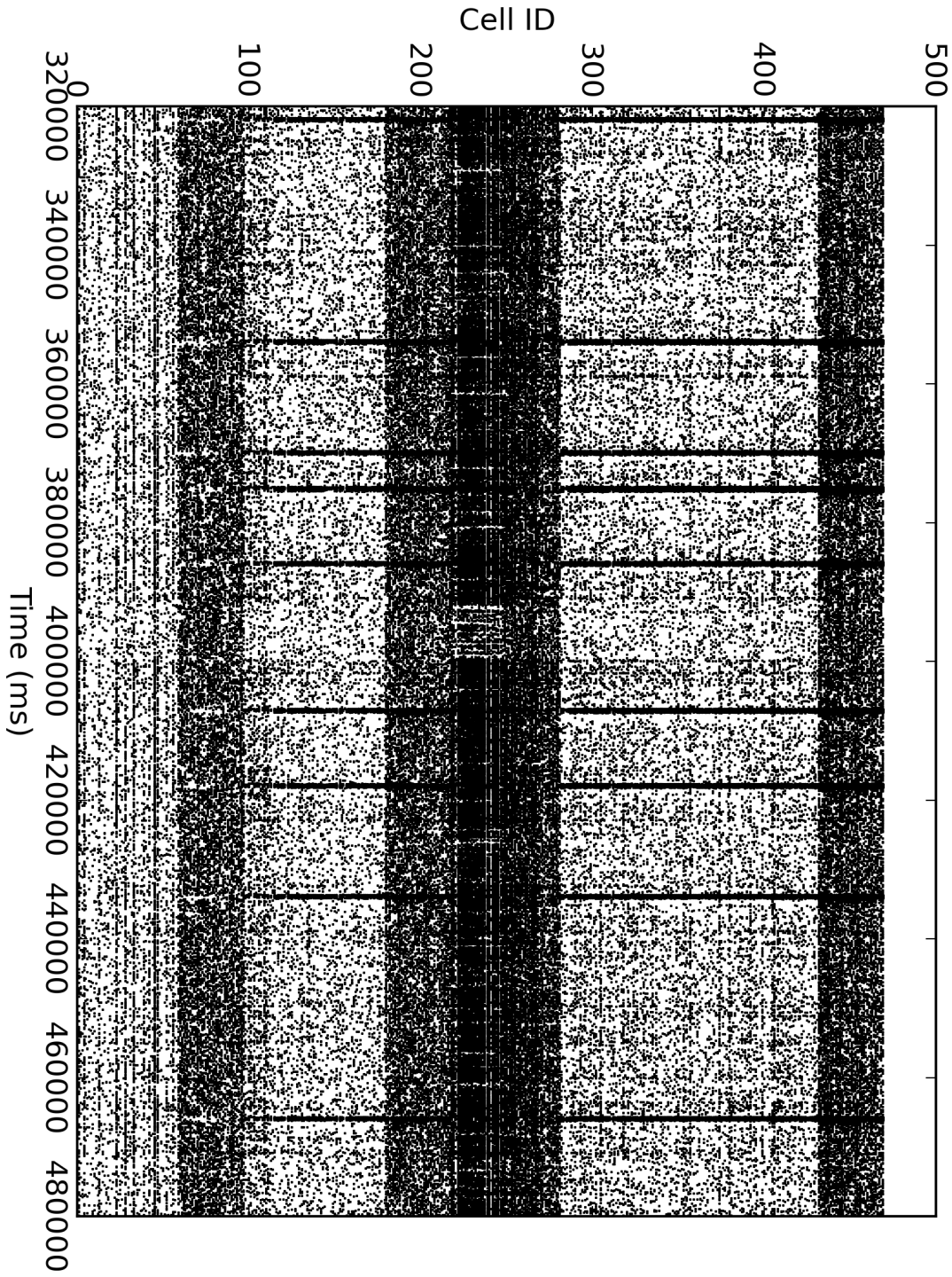
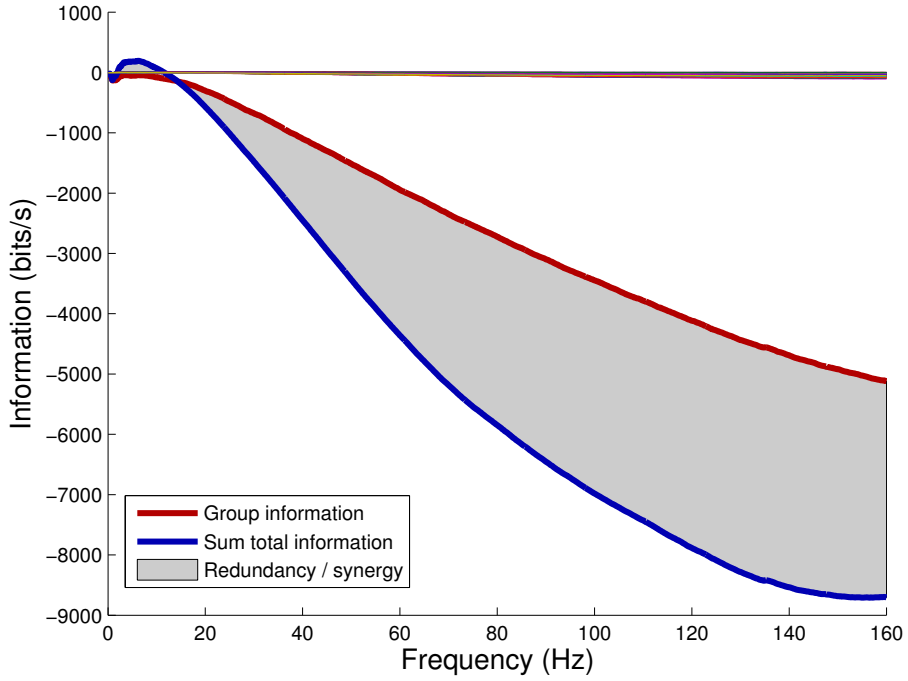
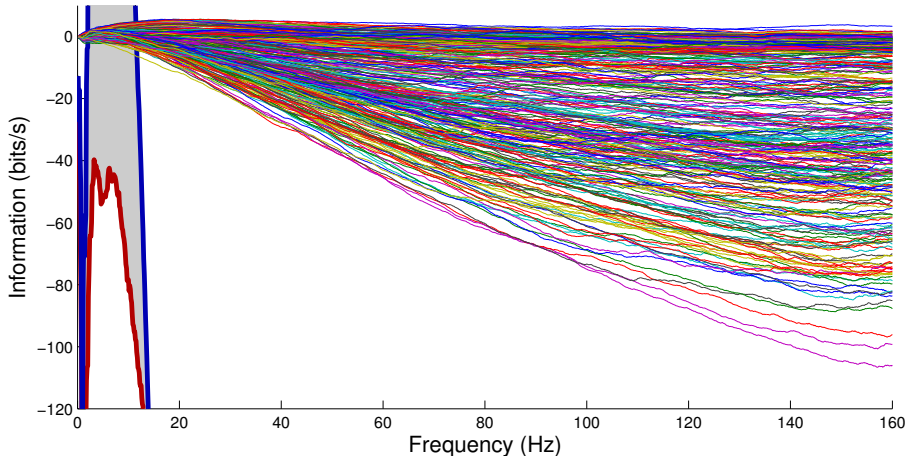


Figure 7.2: Sample raster plot during 10 unique / repeat trials of stimulation of E4 cells with varying frequency. The raster shows increased activity in E4 and varying responses to the different trials, but also displays high-frequency bursting on some trials.



(a) Sum total and group information



(b) Close-up showing individual cell information

Figure 7.3: Sample information graphs from one run using stimulation of E4 cells with varying frequency. The sum total and group information are less than zero, as are most of the individual cell information measures, which indicates that the noise in the calculation was too great to obtain any meaningful information estimation.

low coding rate (representing the proportion of cells active in any one associative state), the percentage of E4 cells stimulated as part of any pattern was reduced to 15% (Abeles et al., 1990). Patterns were generated according to the scheme in section 7.2.4.

Due to the reduced size of the population which was being stimulated, it was necessary to increase the stimulation frequency to 20 Hz to attempt to maintain whole-network firing rates. However, this again increased the tendency of the network towards high-frequency bursting, even at 20 Hz (as seen in figure 7.4).

The results in table 7.2 show that, contrary to intentions, the average firing rate across 'unique' and 'repeat' trials has decreased to little above baseline. The raster plot (figure 7.4) shows very little distinction between the 'unique' and 'repeat' stimuli. One possible reason for this is that random stimulation of only 15% of cells in the network is not sufficient to push the network out of its default basin of attraction, which is the baseline oscillatory rhythm displayed in the rasters. The result of this is that the self-similarity of the network response to the trials, as measured by the coefficient of variation, is low, as is the sum total information, at generally much less than 1 bit/s. The information graphs were qualititatively similar to the previous experiment (figure 7.3), indicating a high level of noise within the calculations.

Associative storage and retrieval of patterns

In order for stimulation of small subsets of cells to be able to affect the response of the whole network, it is necessary for the network to be able to shift into a different basin of attraction depending on the input. By default, the network has only one large basin of attraction (its inherent oscillations), as it is untrained. The initial weights of the network provide possibilities for other basins of attraction, but these must be reinforced and deepened through learning before the network is able to change its state when presented with varying input.

Therefore, a regime of associative pattern storage was undertaken before initiating the retrieval of the stored patterns during the information trials. Before learning, the internal

Firing rate (Hz)		KDE Coefficient of Variation		Info (bits)
Repeats	**Uniques**	**Repeats**	**Uniques**	
0.72 ± 0.31	0.74 ± 0.33	0.3877 ± 0.0226	0.3907 ± 0.0410	0.29 ± 0.51
0.66 ± 0.30	0.68 ± 0.31	0.3851 ± 0.0230	0.3764 ± 0.0262	0.35 ± 0.62
0.66 ± 0.30	0.68 ± 0.34	0.3841 ± 0.0248	0.3845 ± 0.0425	0.19 ± 0.47
0.67 ± 0.30	0.73 ± 0.43	0.3857 ± 0.0281	0.7273 ± 0.2697	0.16 ± 0.47
0.69 ± 0.30	0.70 ± 0.31	0.3837 ± 0.0254	0.3805 ± 0.0247	0.43 ± 0.71
0.65 ± 0.29	0.67 ± 0.31	0.3835 ± 0.0250	0.3765 ± 0.0347	0.41 ± 0.65
0.66 ± 0.29	0.69 ± 0.38	0.3825 ± 0.0235	0.4076 ± 0.0667	0.23 ± 0.53
1.28 ± 0.41	0.68 ± 0.36	4.7065 ± 2.4441	0.4609 ± 0.0943	11.87 ± 8.19
0.67 ± 0.30	0.70 ± 0.33	0.3843 ± 0.0237	0.4054 ± 0.0813	0.33 ± 0.63
0.74 ± 0.31	0.69 ± 0.31	0.5907 ± 0.3969	0.3881 ± 0.0427	0.38 ± 0.62
0.65 ± 0.29	0.70 ± 0.43	0.3846 ± 0.0236	0.5160 ± 0.1314	0.16 ± 0.44
0.68 ± 0.30	0.69 ± 0.32	0.3889 ± 0.0325	0.3875 ± 0.0444	0.24 ± 0.52
0.66 ± 0.30	0.68 ± 0.31	0.3830 ± 0.0247	0.3782 ± 0.0259	0.45 ± 0.68
0.67 ± 0.30	0.69 ± 0.30	0.3826 ± 0.0234	0.3799 ± 0.0237	0.45 ± 0.65
0.67 ± 0.30	0.69 ± 0.31	0.3860 ± 0.0253	0.3770 ± 0.0250	0.32 ± 0.58
0.69 ± 0.30	0.71 ± 0.39	0.3837 ± 0.0232	0.4204 ± 0.0638	0.23 ± 0.61
0.66 ± 0.29	0.70 ± 0.35	0.3832 ± 0.0228	0.5246 ± 0.1630	0.27 ± 0.58
0.66 ± 0.29	0.69 ± 0.36	0.3837 ± 0.0253	0.4465 ± 0.1035	0.22 ± 0.53
0.66 ± 0.29	0.67 ± 0.29	0.3823 ± 0.0240	0.3809 ± 0.0236	0.48 ± 0.69
0.68 ± 0.30	0.70 ± 0.31	0.3838 ± 0.0240	0.3771 ± 0.0301	0.41 ± 0.65
0.70 ± 0.14	0.69 ± 0.02	0.6108 ± 0.9651	0.4243 ± 0.0843	0.89 ± 2.58

Table 7.2: Mean firing rate of E cells over 100 'repeat' and 'unique' trials during stimulation of random patterns, with mean KDE coefficient of variation and mean total information, for each experimental run.

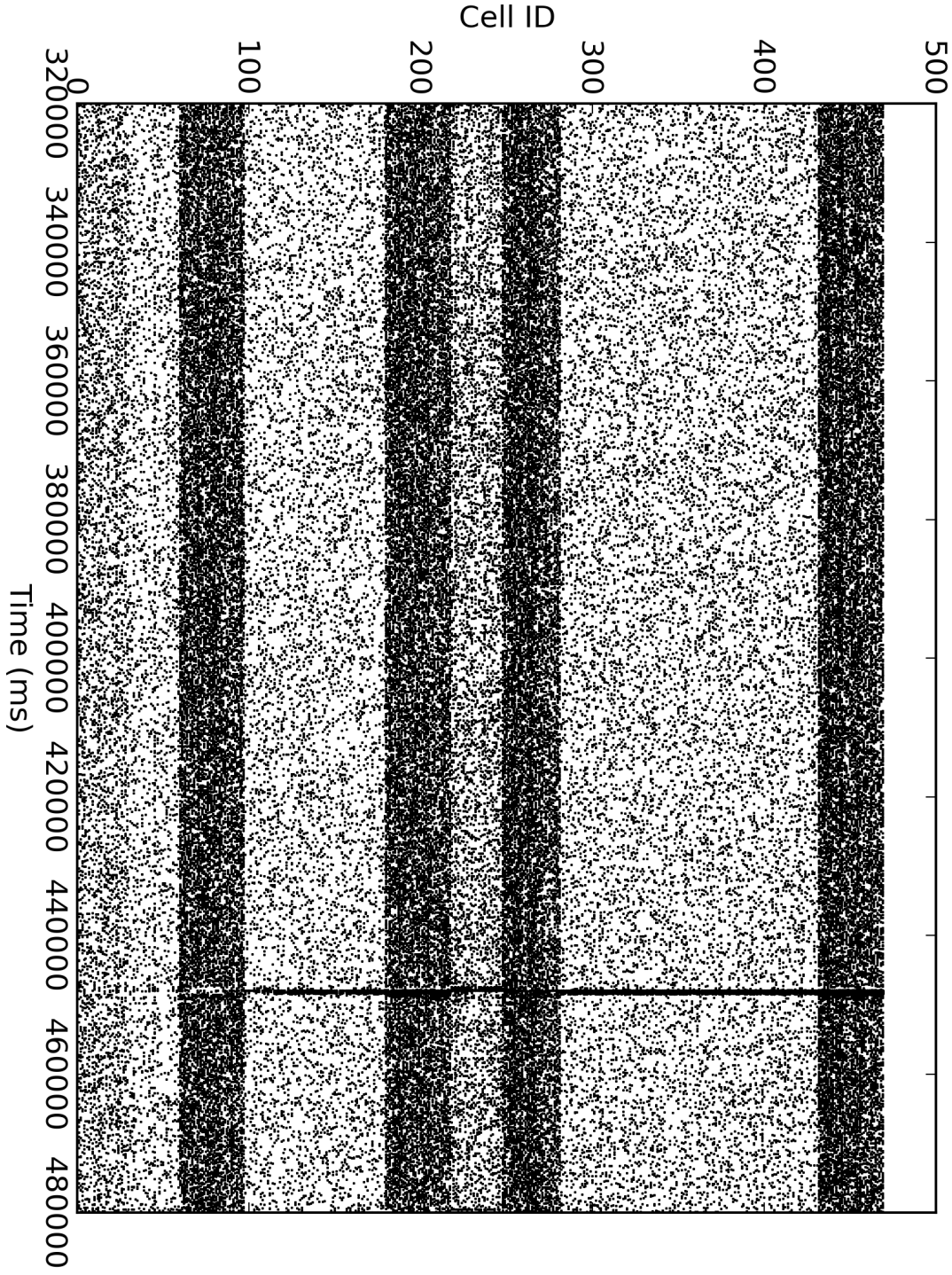


Figure 7.4: Sample raster plot during 10 trials of stimulation of varying patterns of E4 cells with fixed 20 Hz frequency. The raster shows marginally increased activity in E4 but poor variance of responses to the different trials, and also high-frequency bursts on some trials.

weights were reduced to 50% of their baseline values to prevent hyper-potentiation. External inputs were turned off during learning to reduce noise. Patterns consisted of 15% of all E cells within the network (not just E4), to recruit cells from across the network into each pattern. The patterns were presented in random order during learning. Stimulation was provided at very high weight and at a frequency of 12 Hz.

Between each 8 s pattern presentation, there was a further 8 s period of quiescence to allow the network to enter refractory periods and prevent the run-over of the previous firing state to the storage of the next. The network was run at baseline for the first 1600 s to obtain scaling targets, then STDP and synaptic scaling were switched on for 3200 s to learn the patterns. This allowed 200 separate presentations of the 31 patterns chosen at random, or approximately 6.5 presentations per pattern on average, before the retrieval trials began.

During pattern retrieval for the information trials, the external inputs were reactivated. Each 'unique' and 'repeat' pattern was retrieved alternately by low-weight (around neural threshold) stimulation of the relevant E cells in each pattern, with 8 s of quiescence between trials to allow refactory periods for each cell and allow the network to transition smoothly between patterns without "carrying over" activation from a previous trial.

The raster in figure 7.5 shows the network behaviour during learning, with the subsequent behaviour during retrieval in figure 7.6. The behaviour during retrieval shows more well-defined bands of activity than in the previous experiment, but table 7.3 shows that the mean E firing rate for uniques and repeats was barely higher than previously, at just under 1 Hz. The KDE measure actually shows a lower CV value for this regime of pattern storage and retrieval compared to pattern stimulation without storage, and the resulting information measure, at around 1 bit/s, is still very low. This indicates that, despite learning of patterns and an increased firing rate during retrieval, the network was still falling into its default basin of oscillatory attraction.

Firing rate (Hz)		KDE Coefficient of Variation		Info (bits)
Repeats	**Uniques**	**Repeats**	**Uniques**	
1.00 ± 1.09	0.98 ± 1.01	0.4016 ± 0.0365	0.3989 ± 0.0355	1.23 ± 1.59
0.91 ± 0.99	0.89 ± 0.94	0.3980 ± 0.0338	0.3968 ± 0.0449	0.95 ± 1.45
0.92 ± 0.99	0.89 ± 0.94	0.4026 ± 0.0394	0.4001 ± 0.0438	0.91 ± 1.49
0.93 ± 0.97	0.92 ± 0.97	0.4062 ± 0.0500	0.3973 ± 0.0398	1.19 ± 1.72
0.95 ± 1.06	0.93 ± 1.00	0.4054 ± 0.0373	0.4005 ± 0.0457	0.97 ± 1.55
0.92 ± 0.99	0.91 ± 0.93	0.4032 ± 0.0476	0.4021 ± 0.0405	1.06 ± 1.65
0.96 ± 1.05	0.94 ± 0.98	0.3999 ± 0.0332	0.3992 ± 0.0353	0.80 ± 1.41
0.95 ± 1.05	0.91 ± 0.96	0.4029 ± 0.0378	0.3948 ± 0.0453	1.05 ± 1.65
0.95 ± 0.99	0.96 ± 1.00	0.4013 ± 0.0335	0.3947 ± 0.0319	0.98 ± 1.49
0.94 ± 1.05	0.91 ± 0.98	0.4031 ± 0.0313	0.3963 ± 0.0356	1.16 ± 1.95
0.91 ± 0.96	0.91 ± 0.95	0.3996 ± 0.0339	0.4035 ± 0.0348	0.90 ± 1.41
0.93 ± 0.97	0.94 ± 0.96	0.4017 ± 0.0400	0.4006 ± 0.0382	1.03 ± 1.62
0.97 ± 1.05	0.93 ± 0.96	0.3981 ± 0.0375	0.3983 ± 0.0327	0.90 ± 1.45
0.93 ± 1.02	0.90 ± 0.94	0.4055 ± 0.0339	0.4022 ± 0.0462	1.19 ± 1.62
0.98 ± 1.08	0.96 ± 1.02	0.3990 ± 0.0384	0.3984 ± 0.0382	0.97 ± 1.45
0.92 ± 1.01	0.89 ± 0.92	0.4067 ± 0.0369	0.4015 ± 0.0412	0.96 ± 1.62
0.95 ± 1.08	0.92 ± 0.99	0.3969 ± 0.0346	0.4005 ± 0.0360	0.93 ± 1.52
0.93 ± 1.02	0.91 ± 0.96	0.4013 ± 0.0446	0.3975 ± 0.0424	0.83 ± 1.50
0.90 ± 0.93	0.92 ± 0.96	0.4038 ± 0.0429	0.4058 ± 0.0389	1.06 ± 1.64
0.94 ± 1.01	0.92 ± 0.96	0.4045 ± 0.0358	0.4024 ± 0.0466	1.19 ± 1.76
0.94 ± 0.02	0.92 ± 0.02	0.4021 ± 0.0028	0.3996 ± 0.0029	1.01 ± 0.13

Table 7.3: Mean firing rate of E cells over 100 'repeat' and 'unique' trials after associative storage and retrieval of patterns, with mean KDE coefficient of variation and mean total information, for each experimental run.

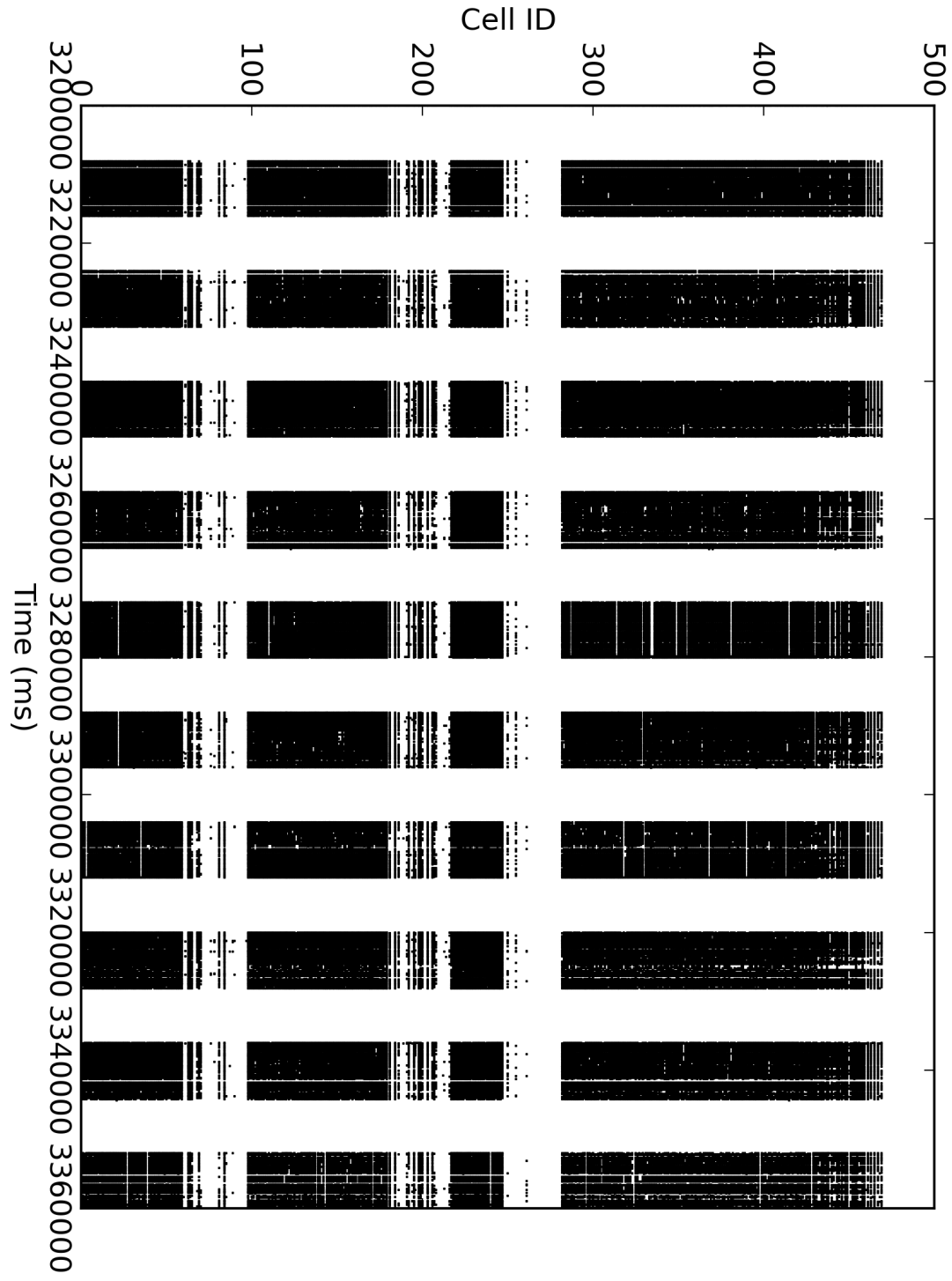


Figure 7.5: Sample raster plot during associative learning of patterns, without external inputs and with 8 s quiescent periods between pattern presentations, showing clear banding of activation.

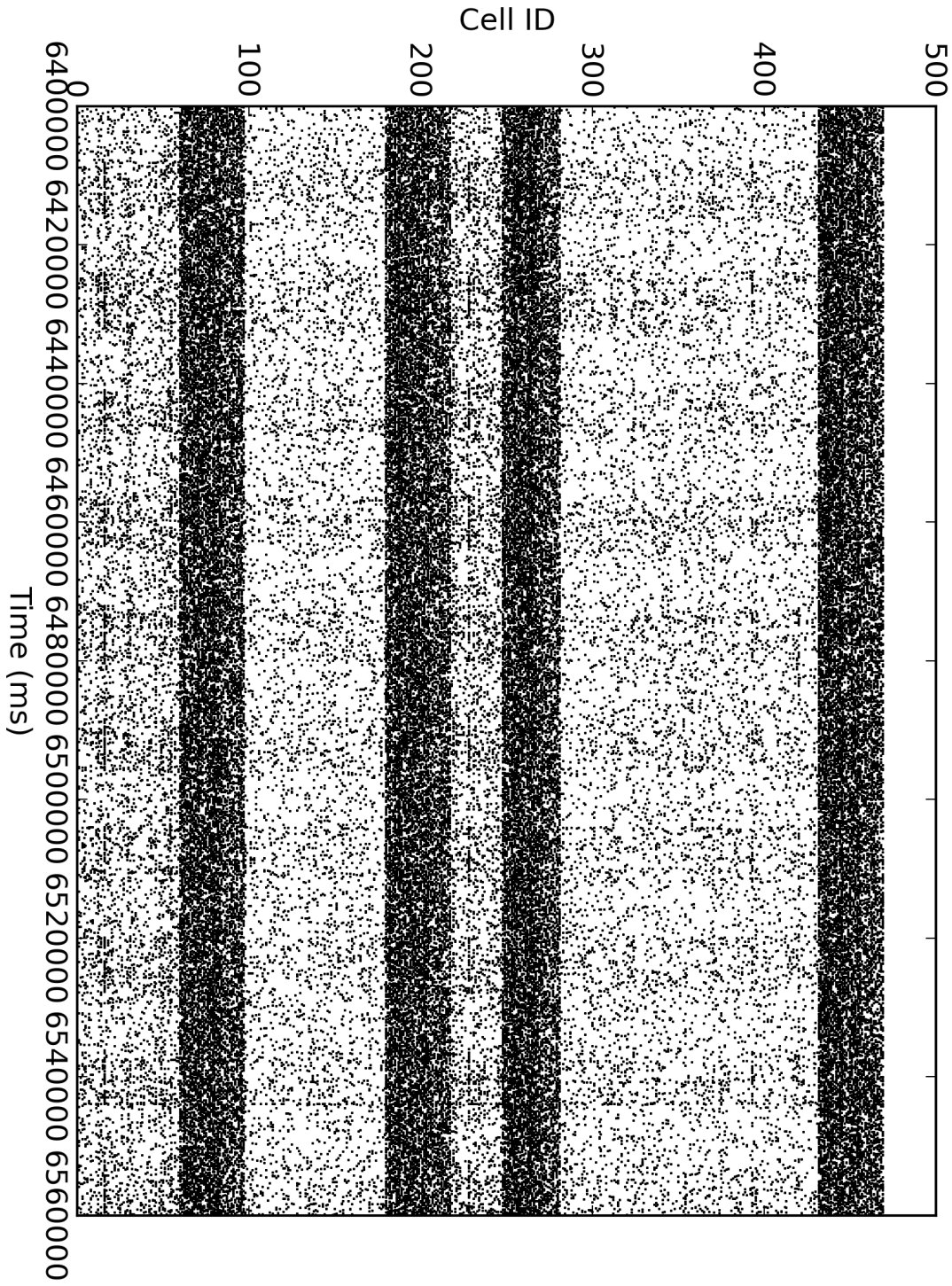


Figure 7.6: Sample raster plot during associative retreival of stored patterns, with 8 s quiescent periods between pattern presentations, showing slight changes in activation between pattern presentations.

Associative storage with clamping

During learning, it can be necessary to prevent recall of previously-stored patterns, as this can simply serve to help reinforce existing patterns rather than forcing the network to learn new ones. If the default basin of attraction of the network is so deep that its dynamics are always retrieved during learning, then it may be necessary to prevent normal network dynamics from occurring during learning. "Clamping" of cells which are not part of the pattern during learning is one way to achieve this (section 7.2.4).

The network was trained in the same manner as in the previous experiment, with cells not part of each pattern held quiescent by above-threshold stimulation of $GABA_A$ synapses. Patterns were learned as before via repeated random presentation, with 8 s periods of quiescence between patterns.

The raster for learning (figure 7.7) shows the effect of clamping during learning, with very clear discrete patterns of cell firing and almost zero firing of other cells during each pattern. However the retrieval raster (figure 7.8) shows no obvious change between pattern retrieval and quiescent states once external inputs are restored. Additionally, the results in table 7.4 show that mean firing rates are only at baseline, with low KDE CV values and a lower mean information value than without clamping, at just over 0.7 bits/s on average.

It is possible that if certain highly-connected 'driver' or 'hub' cells within the network are completely prevented from firing during the learning of most patterns (remembering that each pattern only recruits 15% of cells), then the depotentiation of these cells during STDP essentially removes them from the network and thus erases the dynamics of pattern retrieval which depends on these important cells, thereby leaving only residual activation of each cell from the external inputs.

High-intensity shocking

Due to the problems observed during associative storage and retrieval of patterns, a different approach was attempted. Whilst high-frequency bursting is generally undesirable

Firing rate (Hz)		KDE Coefficient of Variation		Info (bits)
Repeats	**Uniques**	**Repeats**	**Uniques**	
0.60 ± 0.31	0.60 ± 0.35	0.3813 ± 0.0454	0.3874 ± 0.0318	0.71 ± 1.13
0.56 ± 0.30	0.55 ± 0.32	0.3825 ± 0.0397	0.3843 ± 0.0393	0.78 ± 1.14
0.57 ± 0.31	0.56 ± 0.33	0.3786 ± 0.0432	0.3862 ± 0.0332	0.72 ± 1.14
0.56 ± 0.30	0.56 ± 0.33	0.3841 ± 0.0451	0.3783 ± 0.0560	0.63 ± 0.98
0.57 ± 0.31	0.57 ± 0.33	0.3840 ± 0.0551	0.3865 ± 0.0404	0.76 ± 1.21
0.56 ± 0.30	0.55 ± 0.32	0.3791 ± 0.0323	0.3791 ± 0.0425	0.77 ± 1.10
0.57 ± 0.31	0.56 ± 0.33	0.3853 ± 0.0394	0.3830 ± 0.0431	0.78 ± 1.20
0.56 ± 0.30	0.55 ± 0.33	0.3779 ± 0.0411	0.3851 ± 0.0302	0.67 ± 1.08
0.60 ± 0.32	0.59 ± 0.38	0.3752 ± 0.0427	0.3842 ± 0.0454	0.69 ± 1.13
0.57 ± 0.30	0.57 ± 0.33	0.3844 ± 0.0338	0.3822 ± 0.0503	0.66 ± 1.05
0.58 ± 0.31	0.57 ± 0.34	0.3817 ± 0.0404	0.3847 ± 0.0370	0.77 ± 1.12
0.59 ± 0.31	0.57 ± 0.33	0.3846 ± 0.0366	0.3820 ± 0.0438	0.77 ± 1.15
0.60 ± 0.33	0.59 ± 0.33	0.3881 ± 0.0569	0.3847 ± 0.0351	0.79 ± 1.13
0.59 ± 0.33	0.57 ± 0.32	0.4021 ± 0.0880	0.3792 ± 0.0521	0.93 ± 1.27
0.58 ± 0.31	0.58 ± 0.35	0.3782 ± 0.0380	0.3862 ± 0.0286	0.64 ± 1.05
0.57 ± 0.31	0.56 ± 0.33	0.3834 ± 0.0404	0.3887 ± 0.0603	0.70 ± 1.23
0.56 ± 0.30	0.55 ± 0.32	0.3806 ± 0.0382	0.3842 ± 0.0344	0.66 ± 1.04
0.57 ± 0.30	0.56 ± 0.32	0.3810 ± 0.0341	0.3861 ± 0.0332	0.68 ± 1.02
0.57 ± 0.31	0.56 ± 0.33	0.3817 ± 0.0412	0.3834 ± 0.0425	0.69 ± 1.04
0.57 ± 0.30	0.56 ± 0.32	0.3812 ± 0.0324	0.3849 ± 0.0344	0.78 ± 1.24
0.57 ± 0.01	0.57 ± 0.01	0.3828 ± 0.0054	0.3840 ± 0.0028	0.73 ± 0.07

Table 7.4: Mean firing rate of E cells over 100 'repeat' and 'unique' trials after associative storage and retrieval of patterns with cholinergic clamping, with mean KDE coefficient of variation and mean total information, for each experimental run.

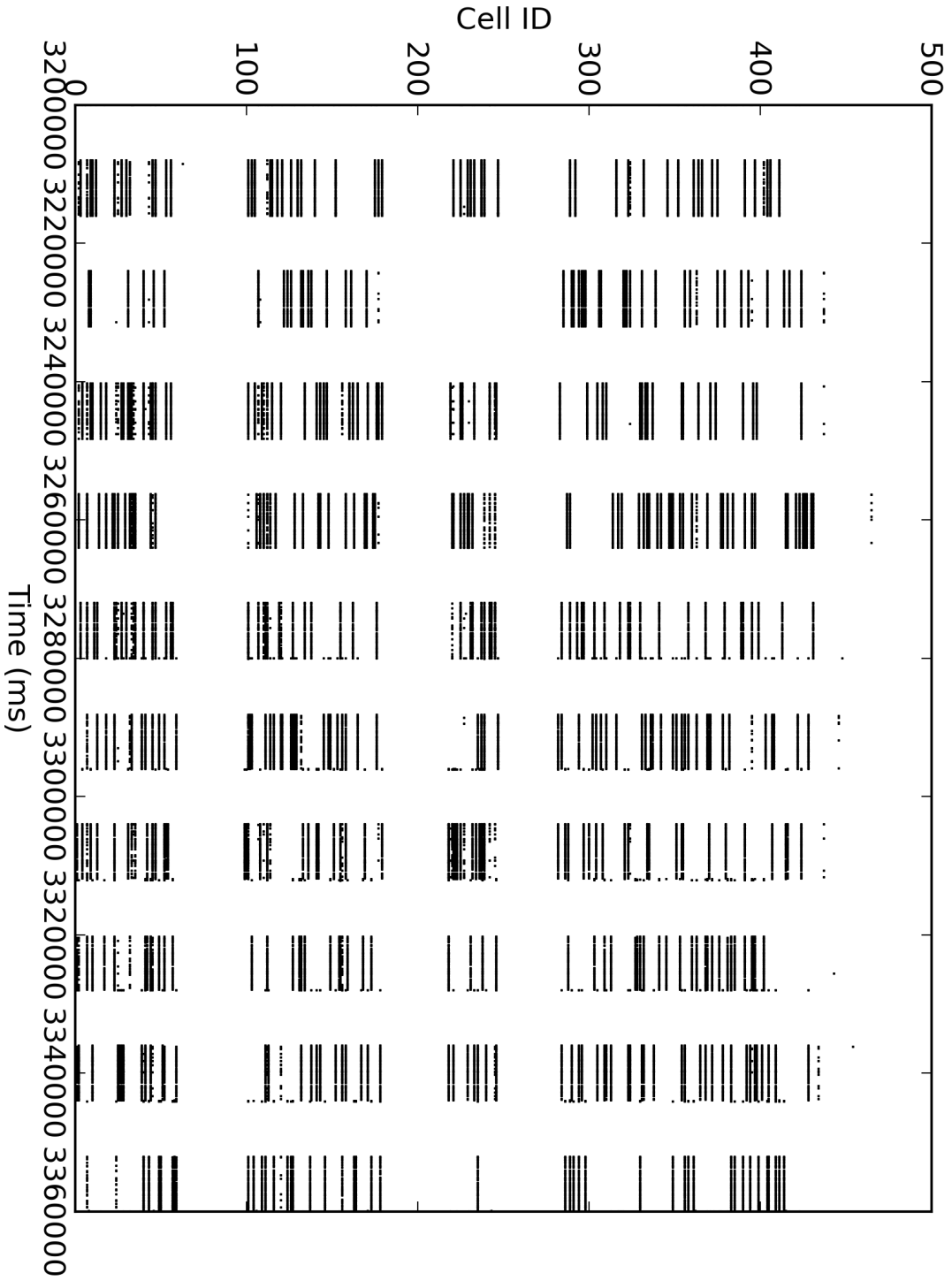


Figure 7.7: Sample raster plot during associative learning of patterns with clamping of non-pattern cells to prevent firing.

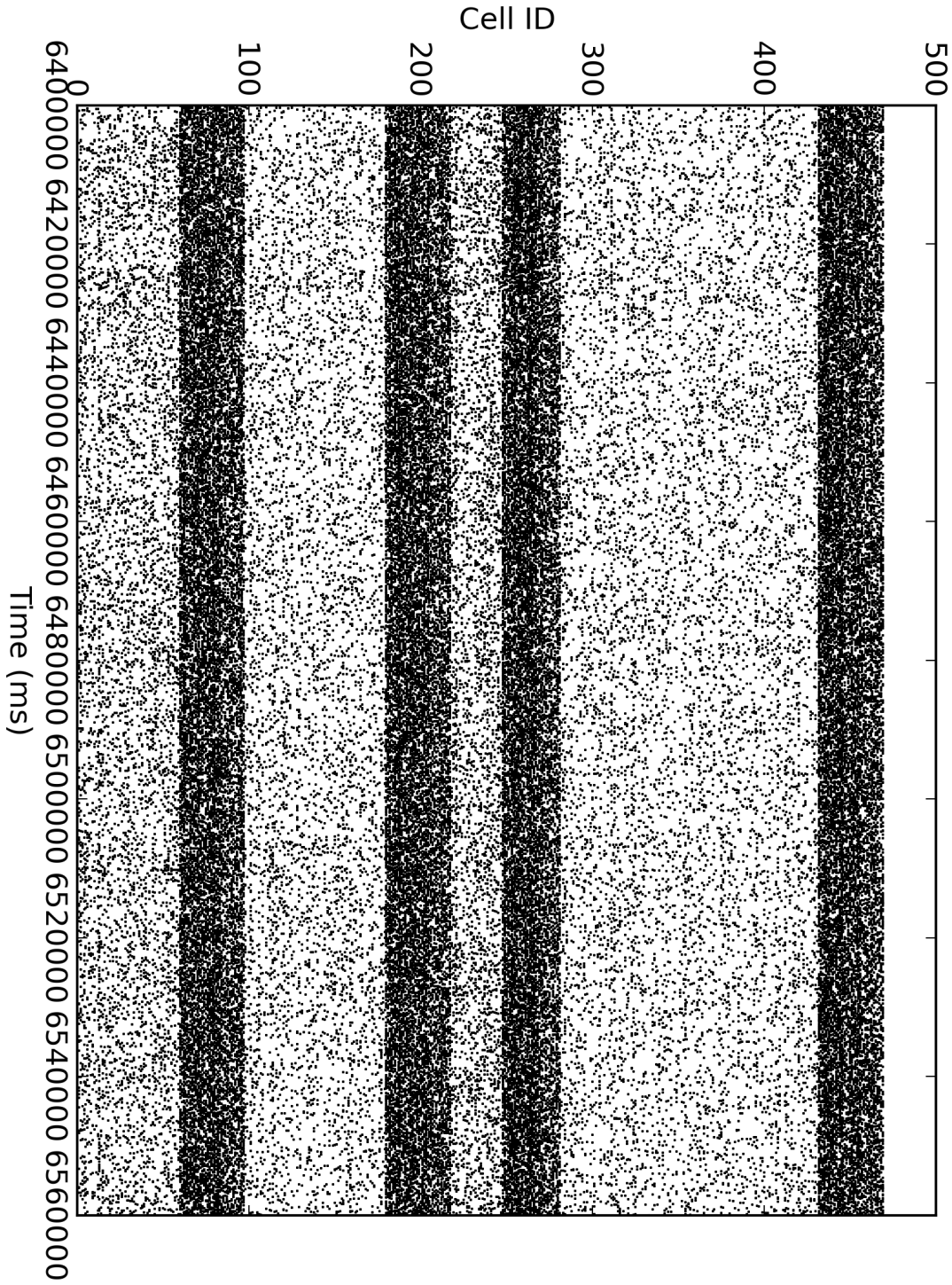


Figure 7.8: Sample raster plot during associative retreival of stored patterns, with 8 s quiescent periods between pattern presentations.

as it renders the network unable to perform computation (Neymotin et al., 2011b), it is possible that short-term shocking of different patterns of cells may allow the network to enter discrete firing regimes based on its intrinsic synaptic wiring, at least for long enough to obtain measures of information contribution before the network transitions to bursting.

To test this, a set of discrete 'unique' patterns and one 'repeat' pattern were created, each consisting of a subset of approximately 15% of all E cells, as in previous experiments. Rather than using STDP to store each pattern before retrieval, the pattern was simply presented to the network at very high super-threshold weight (18× baseline) and at a much lower frequency of 4 Hz, to prevent transition to bursting.

The raster in figure 7.9 shows the effect of this regime of shocking. The mean firing rates of the E cells is now much higher, around 14–15 Hz. The network appears to respond differently to each of the patterns, as seen in the raster, but as evidenced by the KDE CV measure (table 7.5), the self-similarity across all 100 trials is moderate at best. Despite this, however, the mean information contribution of the E cells is much more realistic, at over 16.5 bits/s on average. Transient bursts of synchronised whole-network activity are visible during most of the retrieval trials in the raster, but they are not sustained, and they do not appear to affect the computational abilities of the network to a high degree.

The resulting graph (figure 7.10) shows the desired profile of monotonically increasing sum-total and group information, until it levels off at around 160 Hz. There is synergy (group information greater than sum-total information) up to around 20 Hz, after which redundancy dominates and multiple neurons provide the same information as each other. In the close-up graph, it can be seen that there is a broad distribution of the information-contribution of cells throughout the network, reinforcing the idea that some cells are 'hubs' or 'drivers' of activity, whilst others merely follow the output of other cells and, as a result, have lower or zero information contribution to the network.

A few of these low-information neurons dip to below zero bits/s, which indicates that there is still some noise in the system, but the overall result shows a successful estimation of the information contribution of cells within the network.

Firing rate (Hz)		KDE Coefficient of Variation		Info (bits)
Repeats	**Uniques**	**Repeats**	**Uniques**	
12.70 ± 5.05	14.07 ± 4.62	0.6745 ± 0.1724	0.7958 ± 0.2246	17.03 ± 18.31
14.93 ± 5.81	15.01 ± 5.49	0.6201 ± 0.1401	0.6800 ± 0.1013	3.90 ± 4.14
10.84 ± 4.77	11.27 ± 4.33	0.5891 ± 0.1001	0.5872 ± 0.0810	10.57 ± 10.78
14.49 ± 5.44	15.50 ± 5.33	0.5471 ± 0.0565	0.6723 ± 0.0908	16.97 ± 17.36
25.03 ± 7.15	26.48 ± 6.71	0.5775 ± 0.0704	0.6545 ± 0.1049	20.47 ± 19.31
10.15 ± 5.22	11.95 ± 5.24	0.7197 ± 0.2257	0.6242 ± 0.1498	32.00 ± 31.65
13.54 ± 5.46	14.56 ± 4.91	0.5954 ± 0.1171	0.6299 ± 0.0657	19.79 ± 18.75
15.79 ± 5.89	16.13 ± 5.76	0.6156 ± 0.1104	0.6863 ± 0.1313	4.54 ± 4.37
12.56 ± 5.03	13.49 ± 5.07	0.6728 ± 0.2060	0.7123 ± 0.1870	15.85 ± 13.81
15.25 ± 6.11	16.76 ± 5.53	0.5926 ± 0.0969	0.6824 ± 0.1000	22.36 ± 23.08
15.09 ± 6.15	16.10 ± 5.83	0.6198 ± 0.1546	0.6025 ± 0.0640	16.27 ± 17.11
15.38 ± 5.19	16.82 ± 5.55	0.5782 ± 0.0758	0.8366 ± 0.2832	19.44 ± 21.36
12.59 ± 5.89	14.12 ± 4.96	0.5846 ± 0.0989	0.6268 ± 0.0757	22.07 ± 21.40
11.39 ± 4.61	12.35 ± 4.45	0.6099 ± 0.1086	0.7303 ± 0.1924	16.66 ± 16.18
10.20 ± 4.72	10.63 ± 4.80	0.5288 ± 0.0484	0.5949 ± 0.1195	7.71 ± 9.12
14.21 ± 5.95	15.23 ± 5.37	0.6694 ± 0.1831	0.6514 ± 0.0776	19.70 ± 17.60
17.17 ± 6.40	18.81 ± 5.02	0.6391 ± 0.1624	0.6150 ± 0.1063	20.85 ± 19.30
14.12 ± 6.33	15.49 ± 5.16	0.6268 ± 0.1146	0.5960 ± 0.0708	23.25 ± 19.51
5.50 ± 3.52	6.22 ± 3.30	0.6840 ± 0.1458	0.6707 ± 0.1796	15.82 ± 12.85
18.52 ± 6.68	18.88 ± 5.75	0.6640 ± 0.1996	0.6237 ± 0.0637	7.92 ± 6.67
13.97 ± 3.88	14.99 ± 3.98	0.6205 ± 0.0489	0.6636 ± 0.0659	16.66 ± 6.90

Table 7.5: Mean firing rate of E cells over 100 'repeat' and 'unique' trials during high-intensity shocking, with mean KDE coefficient of variation and mean total information, for each experimental run.

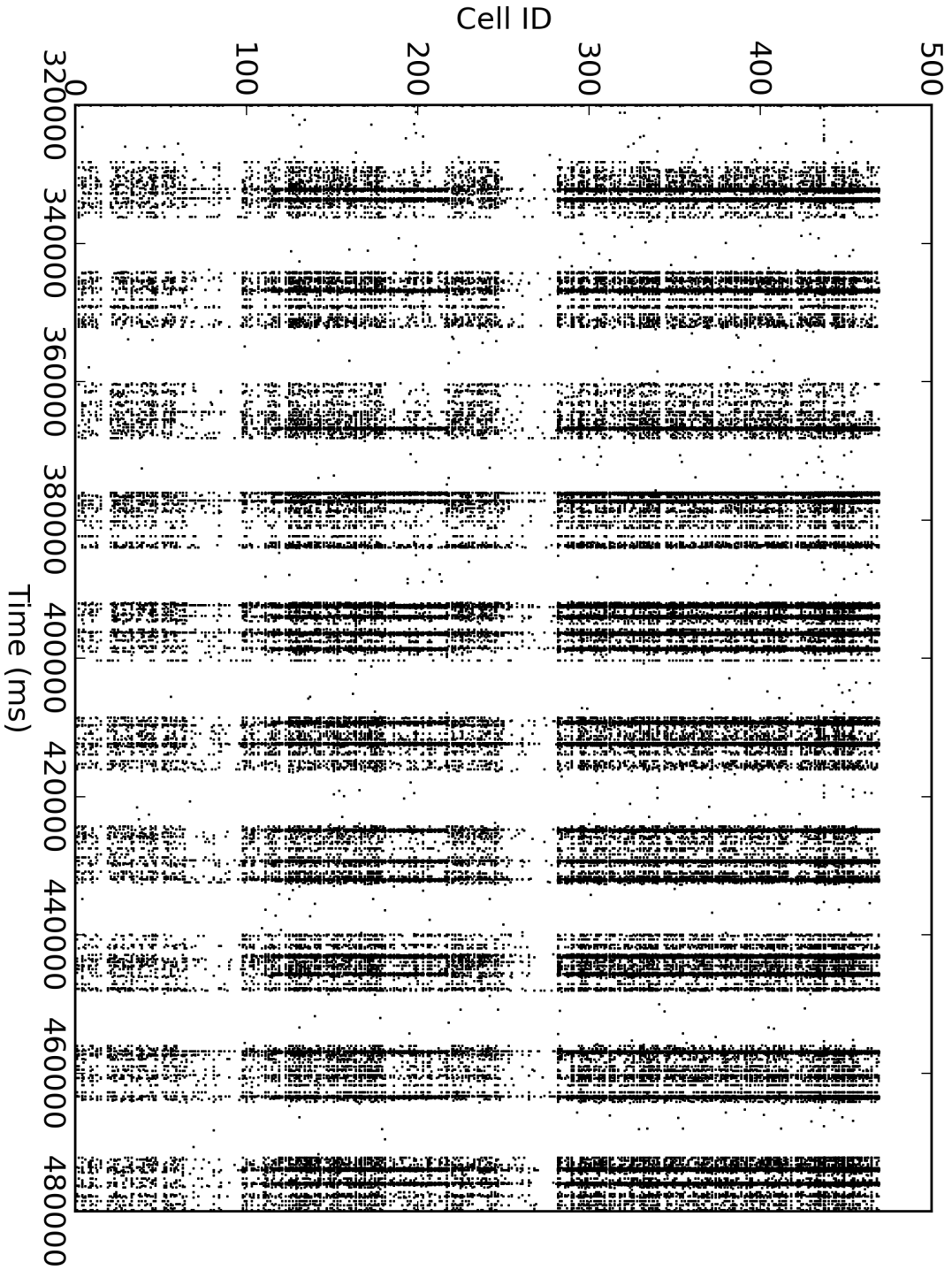
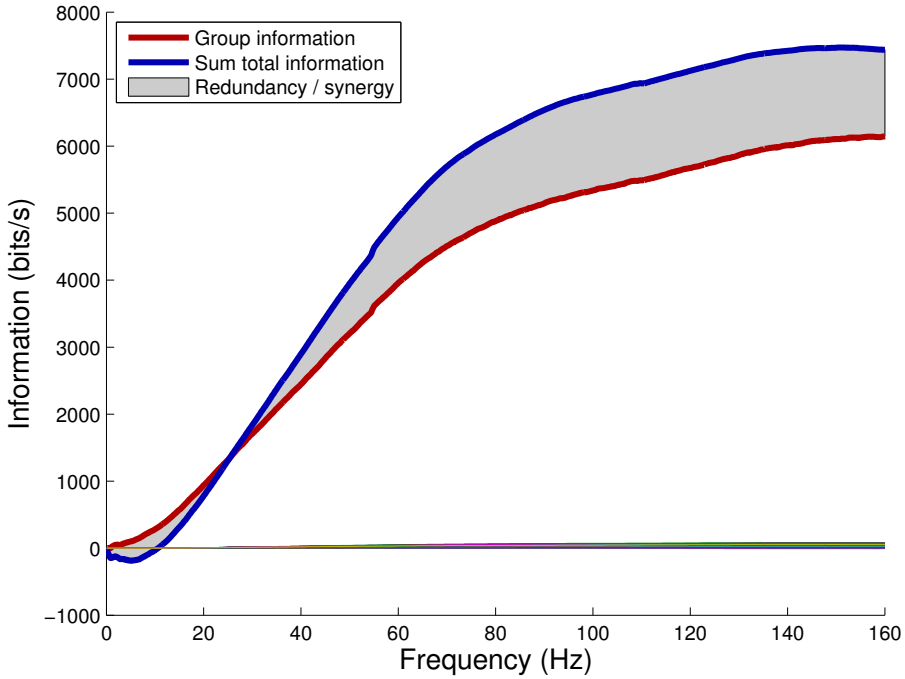
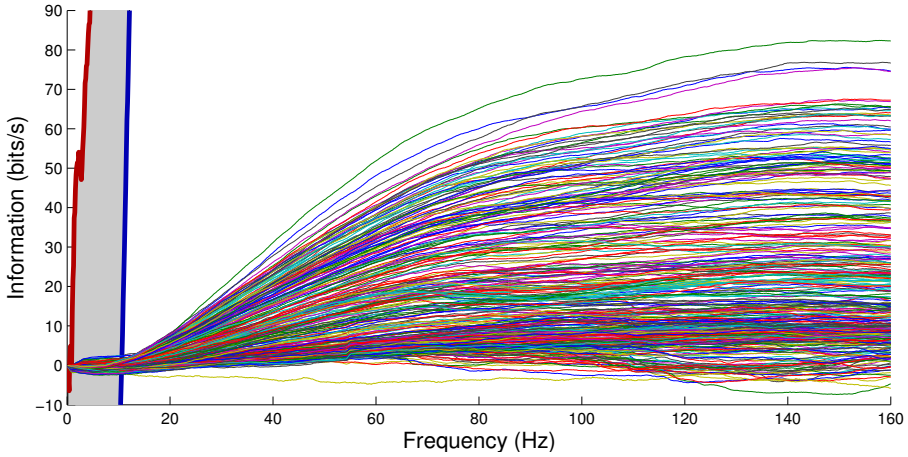


Figure 7.9: Sample raster plot during high-weight shocking of patterns of cells.



(a) Sum total and group information



(b) Close-up showing individual cell information

Figure 7.10: Sample information graphs from one run using high-weight shocking of patterns of cells. The sum total and group information increase monotonically until they level off, indicating that information estimation has been successful. The same general trend can be seen in most of the individual-cell information graphs, too.

7.3.2 Examining information-selectivity

The following experiments were tested over 20 runs with different random seeds on each run for initial synaptic weights, cell positioning, cell wiring, and external input spike trains. For each experiment, mean results and a representative sample run, such that its characteristics were similar to the mean, were displayed, in order to present a clearer view of the data pertaining to an individual simulation.

Each simulation was run for approximately two days of simulated time, but it should be noted that there is no claim to a direct link between the timescales reported in these experiments and real biological timescales. Indeed, progression of Alzheimer's disease is known to take place over the course of months to years, but the two-day simulations are a compromise between biological accuracy of synaptic scaling effects, which take place over "hours to days" (Turrigiano, 2008), and the need for timely presentation of results.

Information contributions of individual cells were obtained according to the method of Crumiller et al. (2011) using high-weight stimulation. Where information contributions are plotted against time of cell death in a scatter plot, the mean of the information values across time was smoothed using a sliding-window average with size 10. At any point of time where more than one cell was deleted within the same discrete time step, the average information value of those cells was used as the representative point for the sliding window.

Information contribution per population

Average information measures were obtained for each cell across 20 different experimental runs using the high-weight stimulation method, and plotted against cell ID to give an overview of the mapping between specific cell populations and their information contribution to the network. Inhibitory cells were not part of the information calculation.

The resulting graph (figure 7.11) shows that E2/3 (layer 2/3 excitatory pyramidal) cells are, on average, the greatest information contributors to the network. Conversely, E6 cells are amongst the least-contributing cells, whilst layers E5a and E5b contain cells with intermediate levels of contribution, with information contribution generally decreasing

from one layer to the next. This is consistent with what would be expected, since E2/3 is the bottom layer of the neocortex, from which signals propagate forwards, whereas E6 is the final layer, which transforms input from lower down in the cortical column but does not directly drive the rest of the network. However, cells within E6 do still contribute information to each other as part of a recurrent network, and there are additionally feedback loops into lower layers which contribute some information back into the network, so the information contribution of E6 is above zero. These results match closely the predictions of Neymotin et al. (2011c), showing the important role of layer E2/3 in driving network dynamics, and the relative high importance of layer E5a.

E4 appears to be a special case, as the sensory layer of the neocortical column via which input from external events is transmitted to the rest of the network. In the absence of an external stimulus, E4 is only driven by the emergent oscillatory dynamics within the network, and does not provide any additional information apart from intrinsic communication via recurrent connectivity.

Random deletion

To provide a baseline against which to compare cell deletion according to AD-related pathology, the model was first lesioned via random deletion of 3 E or I cells every 1600 s, as described in section 7.2.3, and the resulting changes in cell activity and scale factor were plotted. The information contribution of individual cells was obtained and plotted against time of death for each cell.

The results in figure 7.12 show that mean activity was maintained by the synaptic scaling mechanism at around 0.5 Hz (7.12a), declining slightly towards the end of the simulation. The large peak of activity at the start of the graph corresponds to the network response to the high-weight stimulation during the information-probing procedure. Average scaling factors increased exponentially during the progression of deletion (figure 7.12b), but as seen in figure 7.12d, the cells selected for deletion were drawn randomly from either side of the mean line. The information contribution of the selected cells at

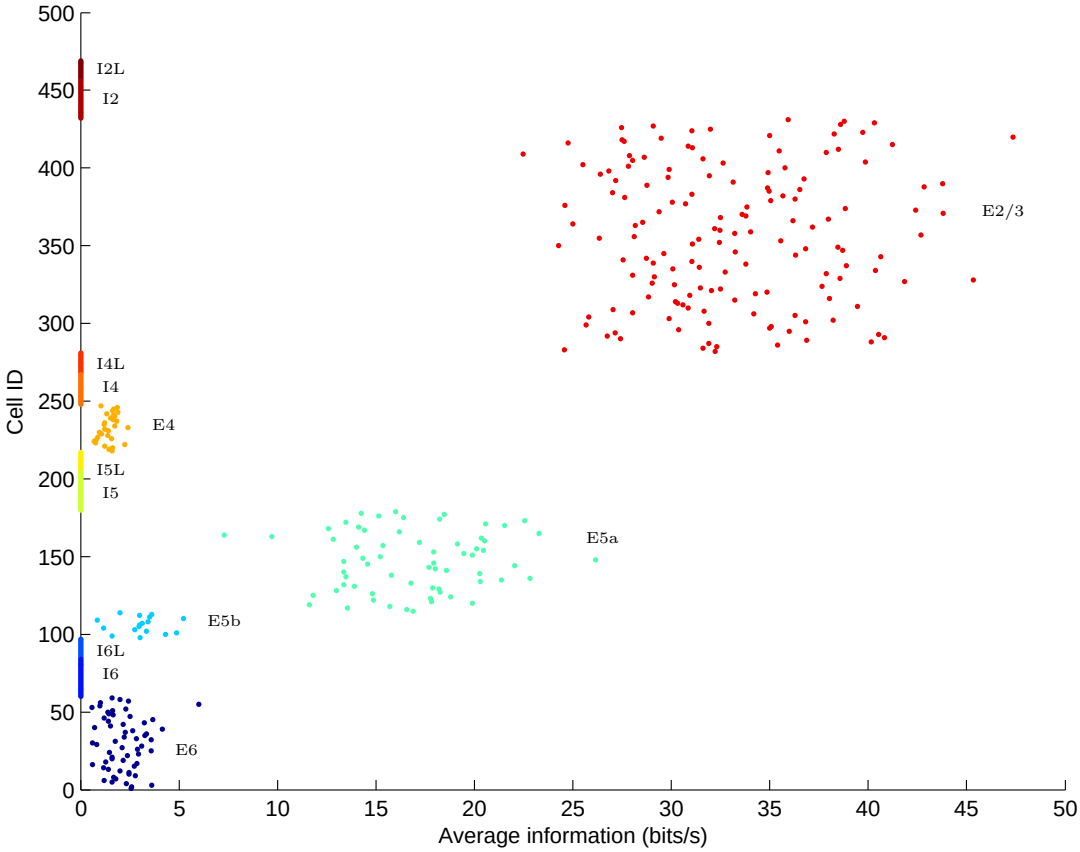
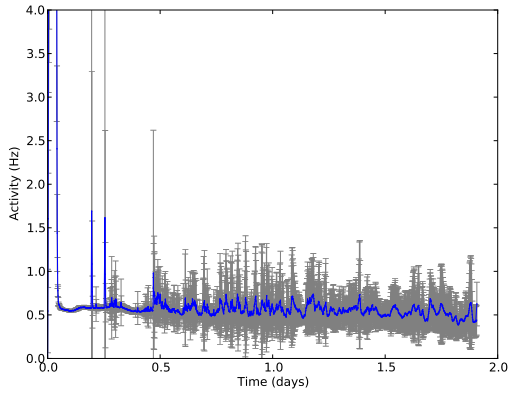
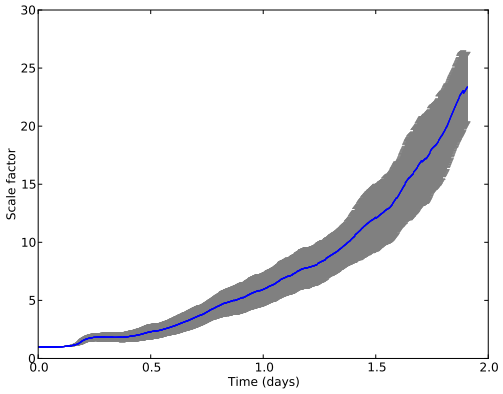


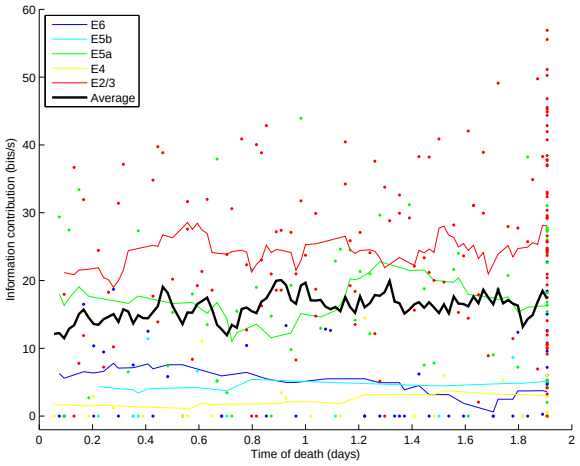
Figure 7.11: Average information per cell over 20 experimental runs.



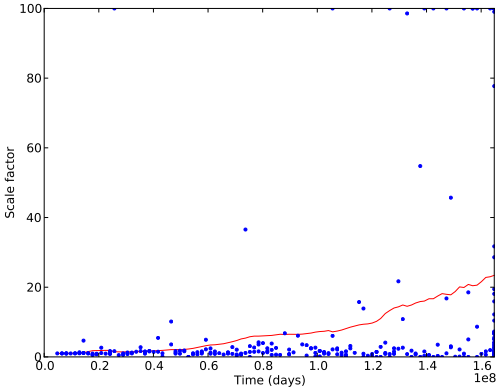
(a) Average E activity



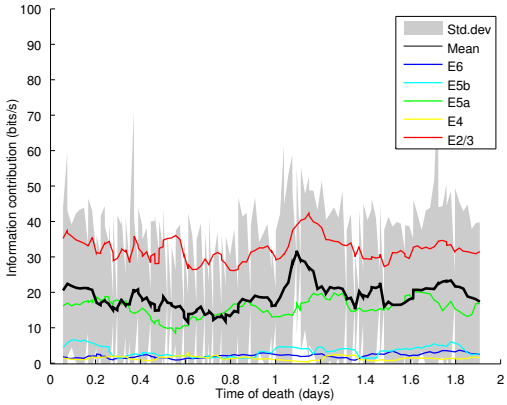
(b) Average scaling factor



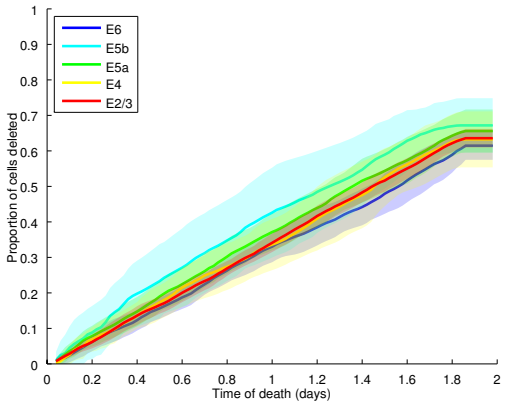
(c) Sample of cell information at time of death



(d) Sample of scaling factors at time of death



(e) Average information of cells at time of death



(f) Deletion rates per population

Figure 7.12: Random deletion (baseline). Mean network activity (7.12a) and scaling factors (7.12b). Sample runs showing individual information contributions of deleted cells against the sliding-window average (7.12c), and individual scaling factors plotted against mean scaling factor (red), as cells were selected for deletion (7.13d). Mean information contribution of populations at time of death (7.12e) and mean deletion rates per population (7.12f).

their time of deletion (figure 7.12e) was generally maintained at around the global mean information contribution value of 16.7 bits/s (table 7.5). A sample from one run is shown in figure 7.12c, again showing a varying average centred around the same value, but with the information contribution of individual cells at the time of deletion (denoted by coloured dot markers) clearly randomly distributed. The uniform random nature of the deletion across cell populations is confirmed in figure 7.12f, which shows highly-similar linear rates of deletion for each of the five populations.

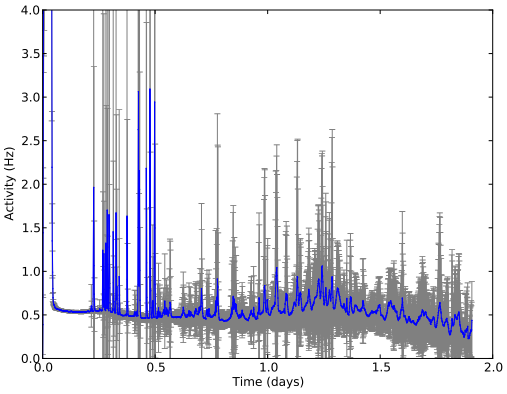
Scaling factor-proportional deletion

The experiment was then repeated with scaling factor-proportional deletion rather than random deletion, as described in section 7.2.3. At each fixed deletion timestep of 1600 s, the three cells with the largest deviations of scaling factor from 1 (either above or below) were selected and removed from the network. The information contribution of individual cells was obtained and plotted against time of death for each cell.

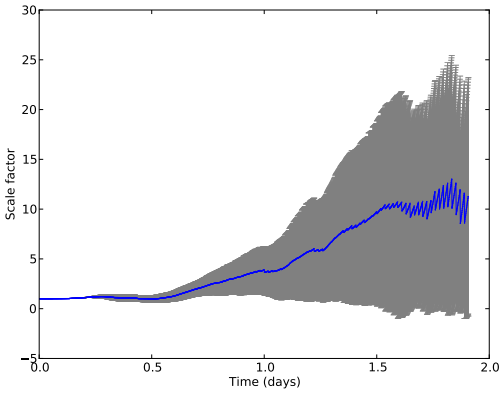
Figure 7.13a shows a slightly different activity profile to baseline random deletion, where activity was again maintained around 0.5 Hz but dropped off more sharply after around 1.25 days. The average scaling factor (figure 7.13b) also shows a different profile to random deletion, with exponential growth at the beginning but a levelling-off of average scaling factor after 1.5 days as the highest-scaling cells were removed at each deletion step.

The scaling factors of the deleted cells in the sample run show that, initially, the fastest scaling was observed in those cells which had to scale up due to under-activity (figure 7.13d). However, after the initial period of deletion, many cells were also forced to scale *down* due to being driven to over-activity by strongly-weighted innervating cells.

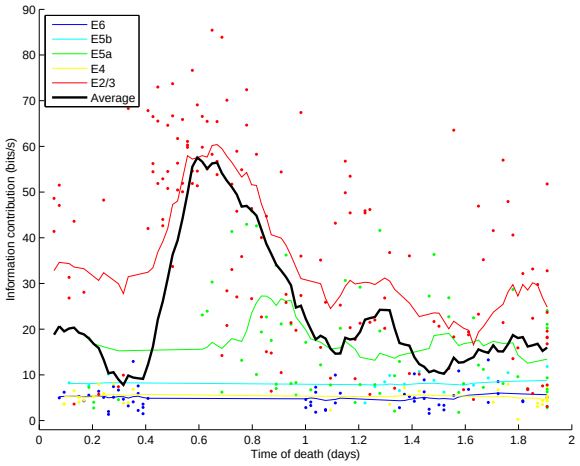
The information contribution in the sample run (figure 7.13c) shows a very different profile of deletion compared to the random baseline. For the very first few rounds of deletion, the selection of cells was essentially random, as evidenced by the averaged spread of the information values of the deleted cells. This was because variations in scaling had



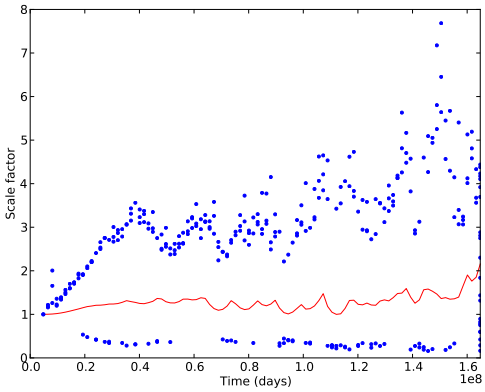
(a) Average E activity



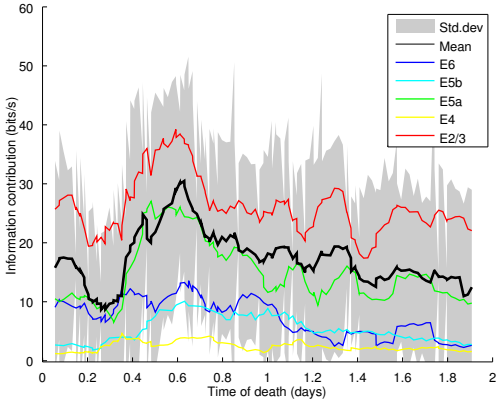
(b) Average scaling factor



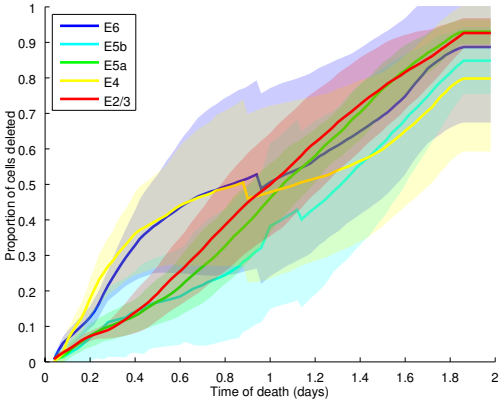
(c) Sample of cell information at time of death



(d) Sample of scaling factors at time of death



(e) Average information of cells at time of death



(f) Deletion rates per population

Figure 7.13: Scaling factor-proportional deletion. Mean network activity (7.13a) and scaling factors (7.13b). Sample runs showing individual information contributions of deleted cells against the sliding-window average (7.13c), and individual scaling factors plotted against mean scaling factor (red), as cells were selected for deletion (7.13d). Mean information contribution of populations at time of death (7.13e) and mean deletion rates per population (7.13f).

not yet propagated throughout the network, as the cells had only just begun to respond to the reduction in activity simultaneously, so scaling factor values were all very similar. However, after this initial brief period of random deletion, there is a clear downwards trend in the average information-contribution of deleted cells, with deletion of low-information cells dominating. In particular, deletion was centred on a large group of low-information neurons from E4 and E6 (the layers with the lowest average information contributions, figure 7.11) for the first 0.4 days. The mean results (figure 7.13e) show a similar profile of deletion across all runs, and the differentiation in deletion across layers is further confirmed in figure 7.13f, in which it can be seen that deletion was fastest in E4 and E6 for the first 0.5 days, although the small but low-information E5b layer actually had a greater proportion of cells surviving until the end.

This period of lowest-information cell deletion in E4 and E6 was followed directly by a period in which the information contribution of the deleted cells was at its highest, with deletion dominated in particular by high-information cells from E2/3, and also the highest-information cells in E5a. The rates of deletion in E2/3 and E5a were approximately linear throughout the experiment, but for a period from 0.4–0.8 days, it was actually the *highest*-information cells from E2/3 and E5a which were selected for deletion.

After around 1.0 days, the rates of deletion per population (figure 7.13f) converged, indicating a more uniform deletion of the remaining cells. It was also shortly after this time that the activity level of the network started to decline, despite compensatory scaling (figure 7.13a). It should be noted that not all of the low-information cells were deleted in the initial phase, as there were a significant number of such cells remaining until the latter stages of deletion.

These results appear to support the hypothesis of information-selectivity via synaptic scaling-driven pathology, but with some important differences to the predictions made in chapter 4. Clearly, when comparing to random baseline deletion, the simulated pathology is affected by the information contribution of neurons. There appears to be a selectivity for low-information cells from across all populations, but particularly E4 and E6, for

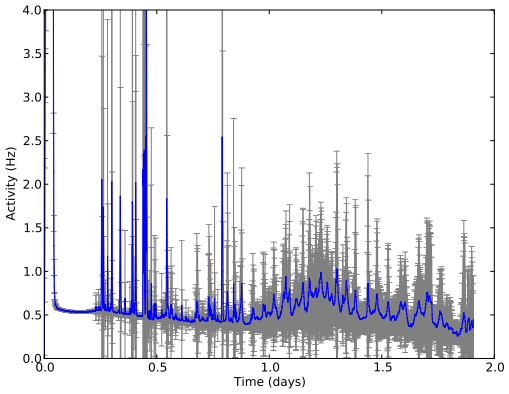
the first period of deletion. However, not all of the low-information cells are deleted in this phase, with those from E5b in particular surviving longer than might be expected. Also, contrary to initial predictions, the low-information deletion phase is closely followed by a period of deletion of the highest-information cells in the network, primarily those from the highly-contributing E2/3 population, but also from E5a, which also has a high information contribution.

Scaling factor-proportional deletion with additional neurotrophic scaling

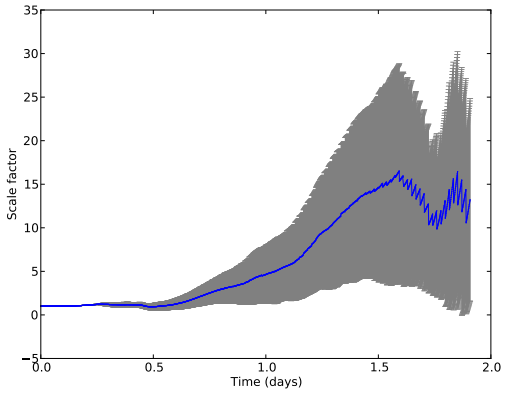
The experiment was then repeated, but with scaling additionally driven via simulated global neurotrophic factors (BDNF and TNF-α). This allowed scaling to be goverened by a global activity signal in addition to the local signal generated by each cell's activity sensor.

The results in figure 7.14 are qualititatively very similar to those in the previous experiment (figure 7.13). The profile of the activity graph (figure 7.14a) shows that the same level of activity was again maintained until around 1.25 days, with a sudden drop-off in the level of activity after this point, despite the global scaling signal attempting to drive the network to maintain its global firing rate. The effects of the global scaling signal can be seen in figure 7.14b, in which the final average scaling value was approximately 50% higher than without the global signal, but this was clearly at a stage of damage beyond the point where additional scaling was able to recover lost activity.

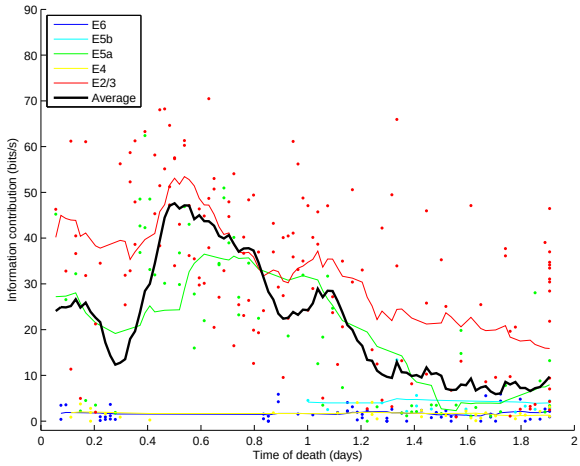
The rate of deletion per population showed a further acceleration of deletion in layers E4 and E6, with the relevant curves in figure 7.14f slightly steeper than those in figure 7.13f, up to 0.4 days. The peak in high-information cell deletion again occurred at around 0.6 days (figure 7.14e), but the peaks of deletion of E2/3 and E5a cells were slightly higher and sharper with neurotrophic scaling than without. Additionally, the period of low-information cell deletion before the high-information peak is marked by a slightly longer and deeper curve than without neurotrophic scaling, indicating that the neurotrophic scaling may be able to increase the information-selectivity effects of the pathology, both



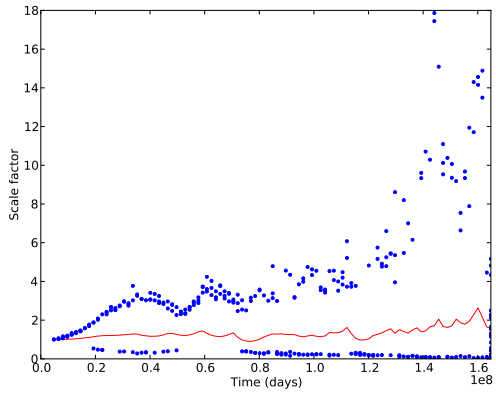
(a) Average E activity



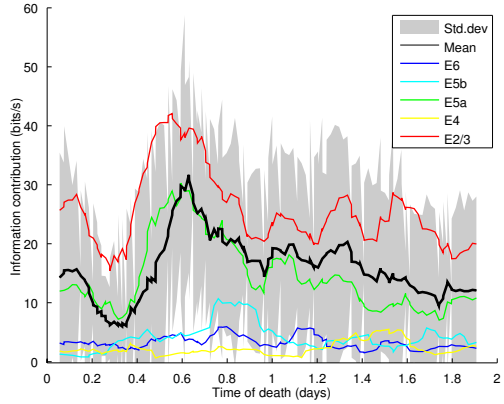
(b) Average scaling factor



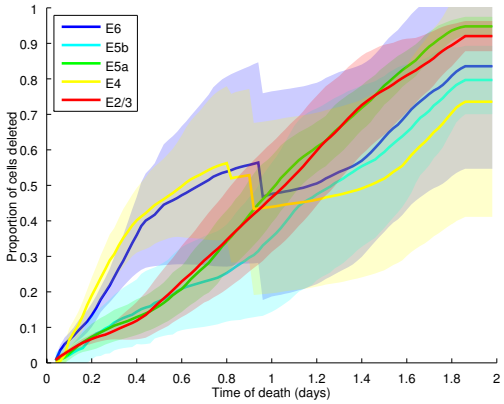
(c) Sample of cell information at time of death



(d) Sample of scaling factors at time of death



(e) Average information of cells at time of death



(f) Deletion rates per population

Figure 7.14: Scaling factor-proportional deletion with neurotrophic scaling. Mean network activity (7.14a) and scaling factors (7.14b). Sample runs showing individual information contributions of deleted cells against the sliding-window average (7.14c), and individual scaling factors plotted against mean scaling factor (red), as cells were selected for deletion (7.14d). Mean information contribution of populations at time of death (7.14e) and mean deletion rates per population (7.14f).

in the low-information deletion phase up to 0.4 days, and in the high-information deletion phase centred around 0.6 days.

Excitotoxicity

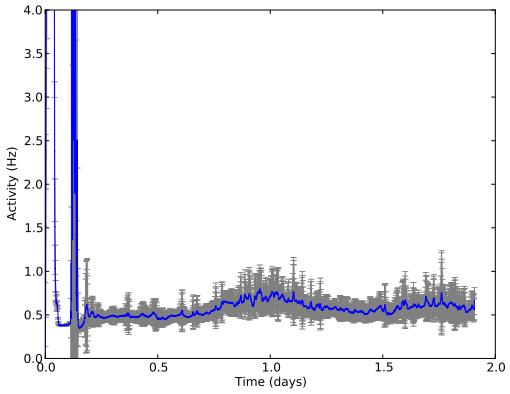
In order to test whether the information-selectivity effects are also observed under more biologically-realistic deletion conditions, the experiments were repeated, but with cell-death determined by excitotoxicity. The time of deletion for each cell was a free parameter, rather than depending on the random or scaling-dependent selection of cells at fixed timesteps.

The results in figure 7.15 were very different to both baseline and scaling factor-proportional deletion. Scaling was able to maintain activity during the deletion process, but figure 7.15d demonstrates that deletion was no longer a regular process, with deletion at irregular intervals. A key feature of the pathology, seen in the activity graph (figure 7.15a), was the large peak in activity at around 0.1 days, lasting approximately 60 seconds following the onset of pathology[1].

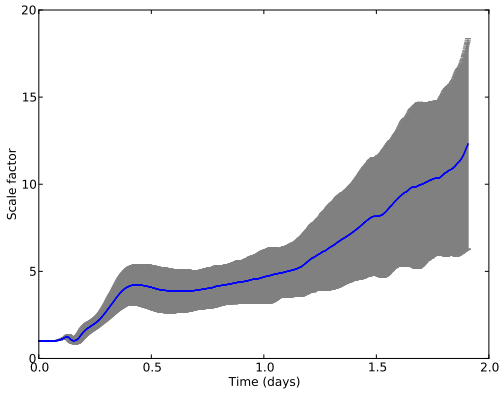
This large peak of activity (actually close to 40 Hz on average, but cropped in the figure) was due to a network-wide period of high-intensity synchronised activity, and was observed in all experimental runs. It is quite different to the transient, short-period (0.5 s) activity bursts observed normally during compensatory synaptic scaling or high-intensity shocking, and responsible for the low-frequency peaks in the activity graph. Comparisons of the two types of bursting can be seen in figure 7.16.

The massive activity peak resulted from instabilities between excitatory and inhibitory populations, magnified by compensatory synaptic scaling. It resulted in the deaths due to over-activity of many cells: between 5–15% of entire populations in one short period (figure 7.15f). However, the early timing of the peak ensured that the deletion was fairly uniform-random in terms of both scaling factor values, which were still around 1 (figure 7.15b), and information contribution values (figure 7.15c). Cell death occurred across the

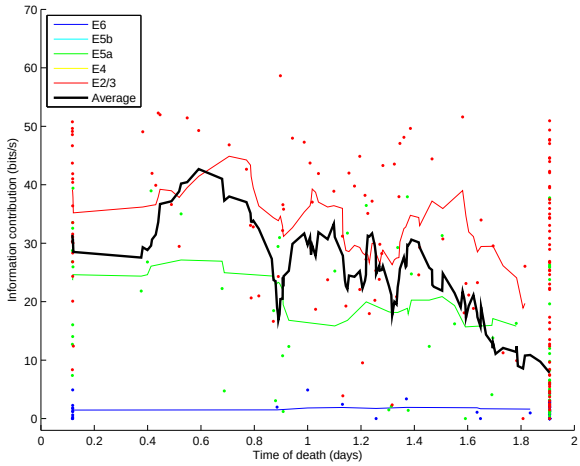
[1]The first peak already in progress at the start of the simulation is due to the high-weight stimulation used for obtaining information contributions of each cell, before the onset of pathology.



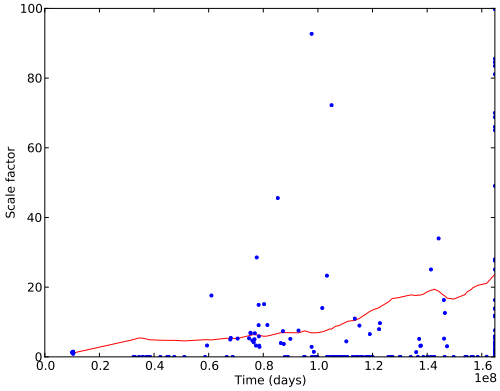
(a) Average E activity



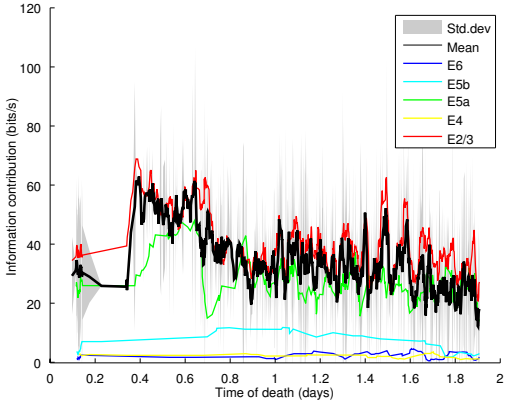
(b) Average scaling factor



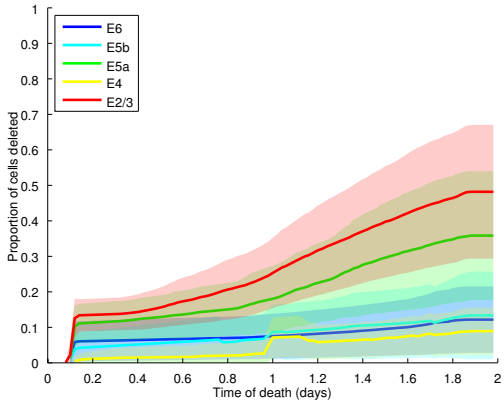
(c) Sample of cell information at time of death



(d) Sample of scaling factors at time of death



(e) Average information of cells at time of death



(f) Deletion rates per population

Figure 7.15: Excitotoxicity-based deletion. Mean network activity (7.15a) and scaling factors (7.15b). Sample runs showing individual information contributions of deleted cells against the sliding-window average (7.15c), and individual scaling factors plotted against mean scaling factor (red), as cells were selected for deletion (7.15d). Mean information contribution of populations at time of death (7.15e) and mean deletion rates per population (7.15f).

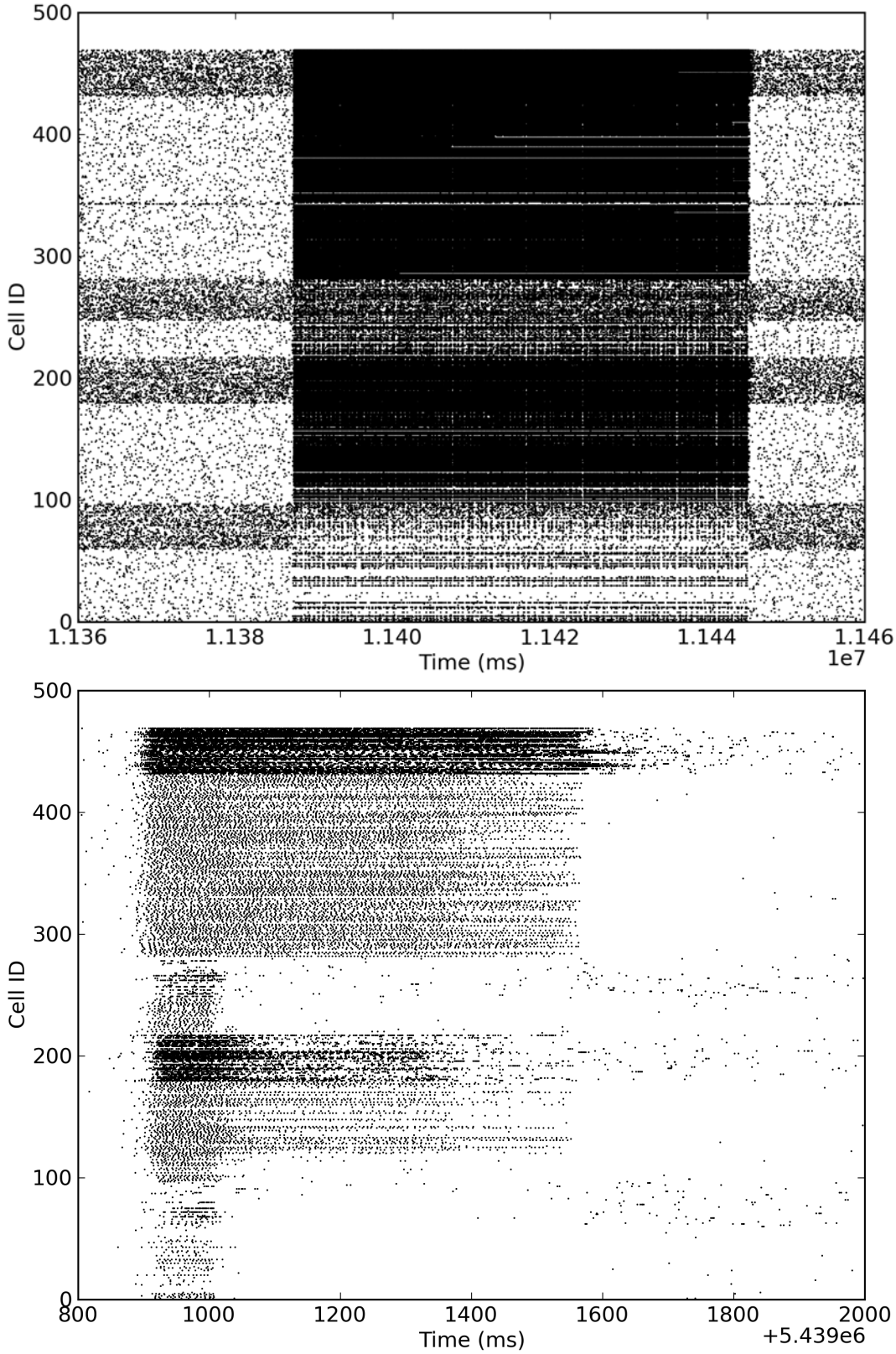


Figure 7.16: Raster plots showing a pathological 60 s burst of high-intensity synchronised activity (top), and a typical transient 0.5 s burst of the sort seen in figure 7.9 (bottom). Each point on the graph marks the time of a spike from the relevant cell.

whole network, but over-activity was particularly strong in layers E2/3 and E5a, with many cells in E4, E5b and E6 conversely spared from over-activity (figure 7.16). This is reflected in the faster rates of cell death in E2/3 and E5a compared with the rest of the network (7.15f).

Following this large activity peak, deletion effectively stopped for around 0.2 days (figure 7.15c) as the large loss in cell populations caused a global reduction in activity below the target level of around 0.5 Hz (figure 7.15a). This caused the rest of the network to scale up in order to compensate (figure 7.15b), and cell death began to progress at a much steadier rate (figure 7.15f). However, deletion remained more dominant in the high-information layers E2/3 and E5a than in E4, E5b, and E6.

The trend in information contribution of individual cells in each population at the time of deletion is a lot less clear than in the previous experiments. However, it appears that the *highest*-information cells in E2/3 were deleted first, followed by the highest-information cells in E5a. This high-information deletion was then followed by a general downwards trend in information-contribution of deleted cells, with the lower-information cells surviving until the end, contrary to the predictions from the previous experiment. However, the data is very noisy due to the abstraction of the deletion criteria from the scaling factor of each neuron. In fact, the observations in figure 7.15d seem to indicate that scaling factor value was not very well correlated with deletion time at all, with a fairly even spread around the mean scaling values for deleted neurons.

Excitotoxicity with additional neurotrophic scaling

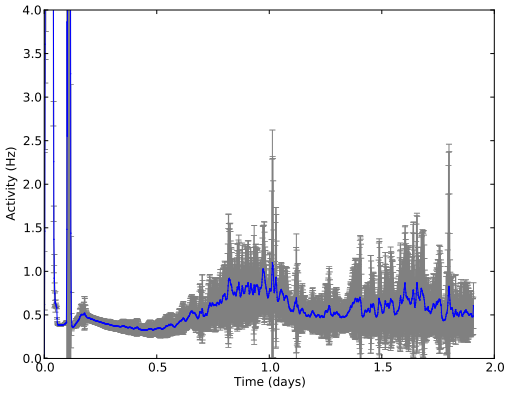
The experiment was then repeated, but with scaling additionally driven via simulated global neurotrophic factors (BDNF and TNF-α). This allowed scaling to be goverened by a global activity signal in addition to the local signal generated by each cell's activity sensor. This emphasis on increasing global firing rates was expected to lead to faster excitotoxic deletion. The deletion rate constant τ_{del} was therefore reduced to one tenth of its previous value, in order to compensate for the increased deletion rate and to allow

deletion to spread across two days of simulation.

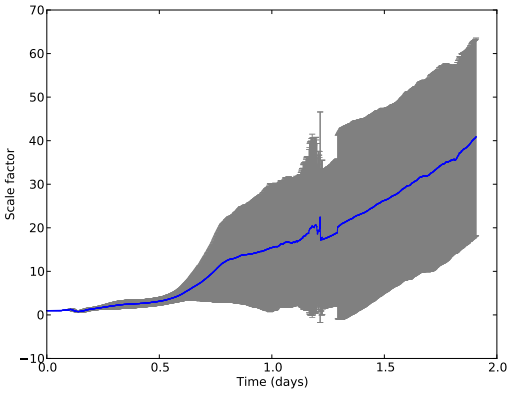
The results in figure 7.17 clearly show the effects of the additional global neurotrophic scaling on global excitation levels, with the activity graph showing large periodic variations in activity (figure 7.17a), with a period of approximately 1 day. There was, again, a long and intense activity peak at around 0.1 days, which corresponds to a large amount of nearly-uniform deletion. However, in this set of experiments the initial deletion was dispoportionately slightly stronger in the low-information E6 population (figure 7.17f), rather than in E2/3 as previously.

Deletion with additional neurotrophic scaling appeared to restore a relationship between scaling factor value and time of deletion (figure 7.17d), but unlike with scaling factor-proportional deletion in which information contribution correlated with time of deletion, the relationship during biophysical deletion was actually reversed. Deletion was dominated by high-information cells in the first period after the activity peak, particularly those in layers E2/3 and E5a (figures 7.17c and 7.17f), with deletion in the latter stages of the simulation confined to low-information cells from E2/3 and E5a, as well as large proportions of the low-information E4, E5b and E6 populations. This downwards trend in information contribution of deleted cells across time is very apparent across all experimental runs (figure 7.17e).

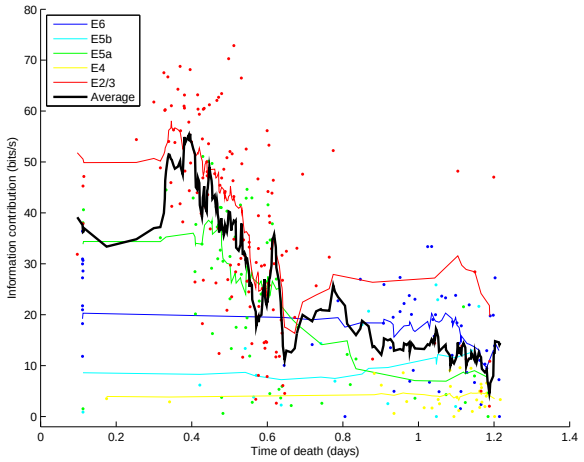
It is interesting to note that the majority of cells selected for deletion on the basis of their over-activity, particularly during the first half of the simulation, were actually scaling *down* in response to being driven to over-activation by neighbouring cells (figure 7.17d). This forced down-scaling due to over-activation appears to correspond with *high* information-contribution. One possible explanation for this effect is that, in an environment where all cells are attempting to maintain firing rates above their 'design' target rate, high-information cells with strong, dense innervation and recurrent connectivity (as found in E2/3 and E5a) are driven disproportionately to over-activation by the stronger feedback within the network. This causes cell death due to the strong feedback loops within the population exacerbating the effects of short transient bursts (as seen in the bottom



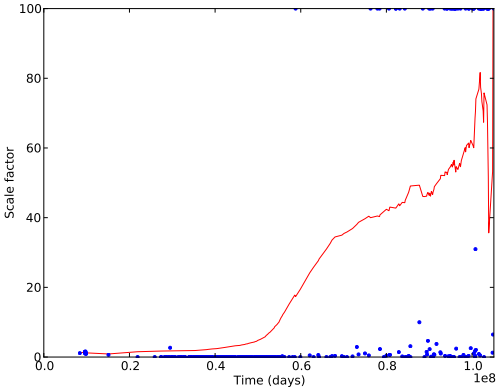
(a) Average E activity



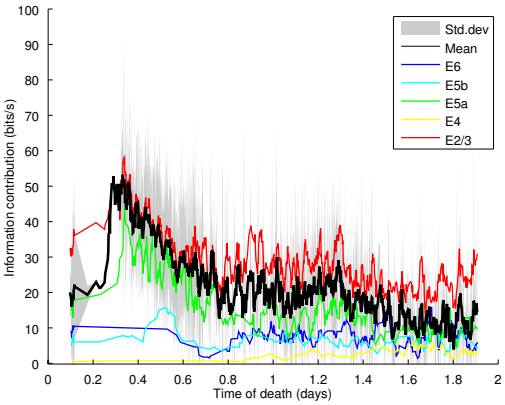
(b) Average scaling factor



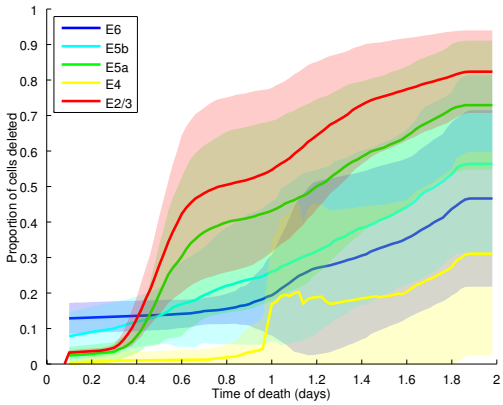
(c) Sample of cell information at time of death



(d) Sample of scaling factors at time of death



(e) Average information of cells at time of death



(f) Deletion rates per population

Figure 7.17: Excitotoxicity with neurotrophic scaling. Mean network activity (7.17a) and scaling factors (7.17b). Sample runs showing individual information contributions of deleted cells against the sliding-window average (7.17c), and individual scaling factors plotted against mean scaling factor (red), as cells were selected for deletion (7.17d). Mean information contribution of populations at time of death (7.17e) and mean deletion rates per population (7.17f).

panel of figure 7.16). Conversely, those cells with weak innervation and low information-contribution are hypoactive, and continue to have low firing rates and fewer instances of transient bursting despite constant up-scaling, due to the lack of positive feedback from within the population and because their high-information 'driver' cells in E2/3 and E5a are constantly dying off. Indeed, the bursts shown in the rasters in figure 7.16 indicate that the low-recurrency, low-information E4 and E6 layers generally had less involvement in network-wide bursting events.

7.4 Discussion

This chapter has investigated the hypothesis of information-selectivity of Alzheimer's disease in a biologically-realistic spiking neural network model of neocortex. This hypothesis states that the cells which contribute the least information to the network should be disproportionately affected by the loss of innervating activation, due to their weaker and/or sparser connectivity. These cells must therefore make greater use of a compensatory synaptic scaling mechanism than their neighbours, in order to recover lost activation levels and return to a target firing rate.

The progression hypothesis of Small (2008) predicts that Alzheimer's disease spreads via the calcium-mediated excitotoxicity of cells which are forced to scale up in response to lost activation from within the network. The combination of Small's hypothesis with the hypothesis of information-selectivity leads to predictions that those cells with the lowest contribution of information to the network are those which must scale up the most, and therefore these cells will die first during the disease pathology. The 'cognitive reserve' of low-information neurons is thus selectively targeted by the pathology, with little or no effect on cognition during the early stages of the disease. This could lead to the delayed onset of cognitive symptoms, and therefore make detection and timely intervention and treatment of the disease much harder.

A recent method by Yu et al. (2010) and Crumiller et al. (2011) for obtaining measures

of mutual information in large networks of spiking neurons was applied to the network, but it was first necessary to design a regime of network stimulation which successfully probes the synaptic weights and reveals levels of information contribution. It was found that the most successful stimulation (as measured by the shape of the resulting group and sum-total information graphs) was high-intensity shocking of different, small subsets of cells across the network, with quiescent refractory periods between each 'pattern' to prevent over-activation spilling across patterns and tipping the network to high-frequency synchronised epileptiform activity. Associative learning of fixed patterns via STDP, including with cholinergic 'clamping' to prevent interference of previously-stored patterns with new patterns during learning, was shown in section 7.3.1 to be unsuccessful as a means of probing the synaptic weights to obtain measures of information contribution.

The information measures were then used to characterise the network in terms of information contribution per population, where each population is defined as a distinct layer of the cortical column architecture. It was revealed that E2/3 and E5a populations were the strongest contributors of information, with E4, E5b, and E6 much less significant drivers of network activity, matching predictions by Neymotin et al. (2011c).

Deletion of cells according to deviation of scaling factor values showed a relationship with information contribution of the deleted cells when compared to baseline. After an initial period of uniform-random deletion as variations in scaling began to propagate, low-information cells, particularly from E4 and E6, were selected for deletion. This selection was based only on the emergent scaling factors of the cells, not directly on the information contribution values, thereby indicating support for the hypothesis of information-selectivity. Scaling with an additional neurotrophin-driven global scaling signal was found to moderately enhance the effect of information-selectivity, with slightly higher rates of deletion of low-information E4 and E6 cells in the first half of simulations, and correspondingly less deletion in the high-information E2/3 and E5a populations.

However, the effects were not as distinct as in the previous experiments which used an abstract Hopfield network model (chapter 4). In particular, the predicted information-

selectivity effect was only observed in the first half of the simulated period of deletion, with the period of deletion of the lowest-information cells followed directly by a period of deletion of the highest-information cells from the strongly-contributing E2/3 and E5a layers. The reason for this switch from low-information selectivity to high-information selectivity during the deletion process is currently not clear.

Small (2008) predicts that levels of synaptic scaling should correlate with cell death, and that cells should in general scale *up* their sensitivity to activation as a response to network damage, in order to maintain average firing rates. This study has shown that, in some cases, significant numbers of cells may actually be forced to scale *down* due to the over-activity of neighbouring cells which are themselves scaling up, and that cells which scale down may also be susceptible to excitotoxic cell death due to the persistent driving over-activation from other cells. Therefore, Small's hypothesis must be updated to incorporate the effect of the *size of deviations* in scaling, rather than purely on the amounts of compensatory up-scaling.

With deletion based on levels of chronic or transient over-activity, rather than on scaling factor, cell death was mostly confined to high-information, *down*-scaling cells in layers E2/3 and E5a, rather than low-information *up*-scaling cells, suggesting that the excitotoxicity-based cell death method employed in this study may be directed more strongly by over-activation as a result of strong recurrent connectivity (indicating high information contribution) in these layers, rather than by compensatory up-scaling of cells in response to hypoactivity in low-information neurons. This is contrary to the predictions of Small (2008). The effect of disease selectivity for high information-contributing neurons was strengthed significantly by the addition of a global neurotrophic scaling signal which aimed to maintain global absolute firing rates by dynamically increasing the target firing rates of individual cells.

It is not yet possible to be certain whether the suggested method of activity-dependent (calcium-mediated) excitotoxicty presented in this study is biologically realistic, despite predictions by Small (2008) and Small (2009). In particular, the highly-replicable presence

of a sustained 60-second period of high-frequency epileptiform activity in the simulation should imply massive neural cell death in response to a prolonged epileptiform seizure during the early stages of Alzheimer's disease. However, this is contrary to clinical observations that cell death remains gradual throughout the disease progression. This massive activity peak is quite different to the transient, short-period activity bursts which are observed more commonly during simulated compensatory synaptic scaling (chapter 6), and related to clinical observations (Busche et al., 2008; Fröhlich et al., 2008; Trasande and Ramirez, 2007), which may be a more reasonable biological mechanism for triggering cell death.

In conclusion, this study has shown that the predictions of Small (2008) should be updated to describe the effects of forced *down*-scaling as well as up-scaling of cells in response to network damage. If the toxicity effects of scaling are abstracted away, and cells are simply selected for deletion based on their levels of scaling devation, then the results of this study appear to support the hypothesis of information-selectivity which was proposed in chapter 4. The cells with the lowest information contribution are the first to succumb to the spread of pathology in the early stages of the disease, thereby possibly masking the presence of the disease via the delayed onset of cognitive symptoms, and making timely intervention more difficult.

A suggested mechanism for excitotoxicity continues to establish a relationship between information contribution of cells and time of death, but reverses the direction of the trend, with the highest-information cells dying first. However, the existence of certain unexpected features of network behaviour during this simulated pathology (specifically, a highly-replicable long-period epileptiform peak early after onset of the disease, resulting in massive cell death) casts doubt on the biological realism of this suggested mechanism. Therefore, it is possible that some other mechanism for excitotoxic cell death exists, which does not display such effects and is more consistent with the observed clinical pathology. More work will be required to identify the precise biological excitotoxicity mechanism, and a suitable computational implementation, but this remains an open ques-

tion in Alzheimer's disease research (Hardy, 2009).

7.5 Chapter summary

In this chapter, the information-selectivity hypothesis of synaptic scaling-driven Alzheimer's disease pathology was tested in a biologically-realistic spiking neural model of neocortex.

Experiments were performed to develop a regime of stimulation which successfully elicits levels of information contribution using the Fourier information method of Crumiller et al. (2011). The resulting information measures were analysed to show that the network displays distinct levels of information contribution in each of its different layers, matching predictions by Neymotin et al. (2011c).

The predictions of Small (2008) were updated to include the observation that some cells actually scale *down* their sensitivity during periods of damage, as a result of strong recurrent connectivity causing feedback loops of compensatory activity to drive these cells to over-activity. The notion of deletion of cells in proportion to their scaling factor was therefore modified to take into account *deviations of scaling factors* from original values, rather than deleting in direct proportion to absolute scaling factor values.

This method of deletion according to deviations in scaling factor, as an abstraction of biological toxicity mechanisms, resulted in a correlation between low information contribution of cells and early cell death in the progress of the disease. These observations were further strengthened by the addition of a global neurotrophic-driven scaling signal, as suggested by Small (2008).

A method for biologically-plausible calcium-mediated excitotoxic cell death was proposed in place of the abstract mechanism based on scaling factor values. Experiments showed that this deletion mechanism also produced a relationship between cell death and information contribution, but there was a clear reversal of the correlation such that high-information cells, rather than low-information cells, were the first to die during the disease progression. This mechanism of cell death appears to be driven by the presence of strong

recurrent connectivity within high-information layers of the cortex, rather than by scaling factor, contrary to the predictions of Small (2008). However, this proposed mechanism for excitotoxicity resulted in an unexplained but fully-replicable period of sudden massive cell death at the onset of disease pathology, which does not match clinical observations, and casts doubt on the realism of this particular proposed excitotoxicity mechanism.

The precise biological mechanism of cell toxicity in Alzheimer's disease has not yet been identified (Hardy, 2009), so it is not possible to validate the computational implementation used in this chapter, but the results using the abstracted scaling factor-proportional deletion mechanism (which are grounded in more general principles than the specific mechanism of excitotoxicity) do appear to support the hypothesis of information-selectivity of synaptic scaling-driven pathology in Alzheimer's disease.

CHAPTER 8

EVALUATION AND FUTURE WORK

8.1 Chapter summary

The aim of this thesis, stated in chapter 1, was "to improve on historical connectionist models of AD by incorporating recent medical knowledge and theories of progression, to better understand the mechanisms of AD at the network level. This should lead to testable predictions, and the possibility of identifying novel treatments to slow or halt the progression of the disease". This chapter evaluates the extent towards which this aim, and the subsequent research questions, were met. A summary of the work undertaken during the course of the project is given, and the contributions to knowledge are reviewed. The work is critically evaluated to show where the methodology used optimal approaches, and where it could be further improved. A method for biologically validating the predictions in the thesis is suggested, and avenues for future work are outlined, including a potential novel treatment for Alzheimer's disease.

8.2 Summary of the work and contributions

In this thesis, a computational modelling approach was used to develop and test the hypothesis of information-selectivity in synaptic scaling-driven progression of Alzheimer's disease.

The background research for this thesis (chapter 2) provided a review of computational models and medical pathologies of Alzheimer's disease, leading to the identification of an abstract Hopfield-type model which displays realistic degradation of memories under Alzheimer's-like atrophy with global compensatory synaptic scaling. This model was then explored in chapter 3, and the effects of properties such as network connectivity and capacity on the progression of Alzheimer's disease were characterised. A positive feedback loop of noisy compensatory synaptic scaling was identified, causing the accelerated degradation of recent memories in Alzheimer's disease, which were themselves preferentially used as drivers of the compensatory mechanism.

In chapter 4, the hypothesis of information-selectivity of scaling-driven pathology in Alzheimer's disease was developed, following predictions by Small (2008) that synaptic scaling mechanisms may direct the spread of disease pathology throughout the network. The information-selectivity hypothesis predicts that scaling-driven pathology is capable of selectively targeting neurons with the lowest contribution of information to the network at early stages of the disease, resulting in a delayed onset of cognitive symptoms. This hypothesis was tested in the abstract network, and the results showed a positive correlation between low information contribution and early time of neural death during scaling-driven pathology.

The significance of these predictions is that the delayed onset of cognitive symptoms makes timely intervention and treatment of the disease much more difficult, as the disease is already well-established by the time cognitive symptoms appear. The information-selectivity hypothesis provides understanding of what may be happening in the early stages of Alzheimer's disease before the onset of cognitive symptoms, and leads to predictions for potential novel treatments later in this chapter.

It was considered important to further test the information-selectivity hypothesis in a biologically-validated realistic model, in order to reduce the effects of abstractions and to enable biologically-testable predictions to be made. Chapter 5 therefore presented a thorough search of the literature for suitable previously-published and validated realistic

spiking models of brain regions which could be used to test these predictions. Candidate models were identified, and consideration was given to problems concerning the implementation and testing of synaptic scaling mechanisms in a biologically-realistic spiking model. This included the problem of determining a neuron's target firing rate without altering network dynamics, and obtaining measures of information contribution between spiking cells in a large network.

Chapter 6 extended the selected biologically-realistic model to incorporate synaptic scaling mechanisms and to allow very long simulation runtimes. The slow-varying activity sensor of van Rossum et al. (2000), which allows a neuron to vary its sensitivity in order to maintain its target firing rate, was modified to work with periodic updates instead of fixed timesteps, making it possible to use the activity sensor with modern fast just-in-time neural simulations rather than being limited to simulations with a fixed timestep. A clear demonstration was then provided of the effects of synaptic scaling on balancing hyperactivity during learning in a biologically-realistic neocortical column model, operating for the first time on realistic timescales of "hours to days". Additionally, the stability of the synaptic scaling mechanism was demonstrated to be disrupted during Alzheimer's-like damage, giving rise to transient epileptiform bursting, matching similar medical observations in AD patients.

In chapter 8, the method of Yu et al. (2010) (summarised in Crumiller et al. (2011)) for obtaining measures of information contribution in spiking neural networks was applied to the model. It was first necessary to develop a regime of network stimulation which successfully probes the synaptic weights, and experiments were performed to demonstrate examples of a successful stimulation regime. The resulting measures of information were then used to demonstrate the differences in information contribution across different layers of the neocortical column, matching predictions by Neymotin et al. (2011c).

The predictions of Small (2008) were modified according to data observed in the model, as a mechanism was identified by which a significant proportion of cells actually scale their sensitivity *down* rather than *up* during compensatory scaling processes. The model was

then lesioned in various ways, revealing the existence of a relationship between information contribution and cell death during synaptic scaling-driven pathology in simulated Alzheimer's disease, compared to baseline random cell death.

This relationship was then further explored, showing that an abstracted toxicity mechanism which causes cell death in proprtion to scaling factor values, as predicted by Small (2008) and Small (2009), is capable of selectively targeting cells with low information-contribution to the network in a biologically-realistic computational model. This provides a mechanism for understanding the delayed onset of cognitive symptoms in AD, which makes timely intervention and treatment more difficult.

A further experiment was performed in which the most immediately intuitive mechanism for excitotoxicity was tested, using chronic over-activation of cells rather than abstract scaling factor-proportional cell death. The results maintained a relationship between information contribution and cell death time, but showed a complete reversal of the trend, with high-information cells dying first in the early stages of the disease. The addition of a global neurotrophic-driven scaling signal further strengthened the results, both in this case and in the abstract scaling factor-proportional cell death case. However, certain features of this proposed excitotoxicity mechanism suggested that it may not be biologically realistic, including the replicable presence of a massive period of uniform-random cell death at the onset of the disease pathology.

In conclusion, this thesis has developed and tested the hypothesis of information-selectivity of synaptic scaling-driven pathology in Alzheimer's disease. Testing of the predictions in both an abstract Hopfield-type computational model, and in a biologically-validated realistic spiking neural model, under conditions of cell death based on existing predictions of toxicity related to synaptic scaling mechanisms, produced results which support the hypothesis, although specific excitotoxicity mechanisms proved harder to identify. This work therefore provides a mechanistic explanation for the delayed onset of cognitive symptoms in Alzheimer's disease, and provides a model from which biologically-testable predictions and potential novel treatments may be identified.

8.3 Critical evaluation

> "In the same way that all roads lead to Rome, there may be multiple different ways to build a network with similar output properties [...] Different combinations of intrinsic neuronal properties as well as connectivity patterns can lead to the same properties at the network level [...] In other words, quite distinct combinations of parameters can lead to the same firing regime" – Anderson and Kreiman (2011).

The principles underlying this quotation are very important. It must be remembered at all times that just because an implementation of a model (no matter how detailed) appears to work as expected, it is not necessarily biologically accurate. Indeed, the addition of extra complexity to a model results in the creation of extra parameters, further enhancing the potential solution space of "correctly-working" models.

Ultimately, all computational models are abstractions of biological processes at some level; the only completely accurate model of the brain is the brain itself, but modelling allows researchers to simplify the processes involved, and to use the resulting systems to generate testable hypotheses based on repeatable data. As stated in the introduction, abstraction is both necessary and desirable as a way to understand data, as long as such abstractions are made with foresight and justification. For this reason, care was taken throughout the research to incorporate biological realism into the models where possible, with clear identification and justification of abstractions where they were necessary.

Despite these guiding intentions, there are some areas of the work in this thesis which are possibly too highly abstracted, and in which addition of further biological realism may reveal new or different results. The first of these areas relates to the focus throughout the research (with the exception of the section on tau pathology in chapter 3) on the deletion of neurons rather than synapses. Alzheimer's disease involves both neuronal death and synaptic dysfunction, and the results of the tau pathology experiments showed that synaptic suppression can yield very different profiles of damage when applied to the network. It is not currently known to what extent the addition of specific tau-driven synaptic dysfunction to the model would alter the predictions of the information-selectivity hy-

pothesis. However, the synaptic scaling-driven pathology, which formed the basis for the information-selectivity hypothesis, operates only on the basis of reductions in synaptic input to neurons, and is agnostic to the cause of these reductions (whether due to synaptic dysfunction or neuronal atrophy), so the differences in behaviour may not be major if tau-driven synaptic dysfunction were to be added to the model.

Another area in which the models were perhaps over-simplified is the lack of additional plasticity during the pathological phase. It is conceivable that the addition of STDP or other learning rules which alter synaptic weights in competition with synaptic scaling, or even the generation of new synaptic connections between remaining neurons, would direct network damage to new, previously-unpredicted areas. These effects were not considered in the pathological predictions of Small (2008) either, so an investigation into this aspect of the pathology could be both productive and insightful.

Spiking neural networks in the brain consist of complex interactions between excitatory and inhibitory cell populations. As evidenced in chapter 6, inhibition is vital for the correct operation of such networks, and the disruption of the inhibitory-excitatory balance (for example, through runaway hyperpotentiation during learning without synaptic scaling) leads to catastrophic changes in network behaviour. However, this thesis examines information contribution and subsequent information-selectivity only in excitatory cell populations in the spiking neural model. Whilst this is partially justifiable, as inhibitory interneurons are not usually considered to be a direct part of computational processes in the brain, inhibitory populations still play an important role in governing excitatory cell behaviour. Indeed, some experiments show that death of inhibitory cells during excitotoxicity may precede death of excitatory cells, thereby further exacerbating the excitotoxicity effects (Iglesias and Villa, 2006), so the information contribution and susceptibility to death of inhibitory cells warrants further investigation.

Associative neural networks, including those found in the brain, are structured in such a way that they are capable of storing specific patterns of activation as basins of attraction within the synaptic weight space. Such networks constantly transition from one

learned pattern to another, depending on the input to the network. Therefore, associative networks are highly likely to be in an attractor state at any given time, rather than in a uniformly active state. However, the spiking model used in this work had only one very deep basin of attraction, which defined its default uniformly active oscillatory behaviour. If the spiking model could be extended to successfully store and transition between a variety of attractor states, this would provide opportunities to examine the different distributions of neuronal activity and the effects that changing between stored states may have on the progression of damage. It would also enable quantitative measures of network performance during recall of stored patterns, allowing direct comparisons with the abstract model examined in chapter 4 which are not currently possible.

Finally, although an abstract cell death mechanism grounded in the principles of synaptic scaling-driven excitotoxicity has been shown in this thesis to selectively target neurons with low contribution of information to the network at early stages of the disease, the most obvious biological mechanism for excitotoxicity showed some replicable biologically-implausible results, as discussed in chapter 7. The precise mechanism of excitotoxicity in Alzheimer's disease is still not known (Hardy, 2009), so validation of proposed computational mechanisms is naturally difficult. The development of a more plausible mechanism for excitotoxic cell death would be beneficial, as it would enable the generation of further biologically-testable predictions, including possible bio-markers which may reveal the presence of the disease via externally-visible signs before the onset of cognitive symptoms, such as differences in EEG scans (Bhattacharya et al., 2011).

8.4 Future work

As discussed in the previous section, the model could benefit from increased biological realism, and could potentially reveal different patterns of network damage, with the addition of various features. These potential enhancements include adding tau-driven synaptic dysfunction to the existing neuronal atrophy, and the inclusion of learning and synaptic

generation during the progression of damage. Further experiments focussing on the role of inhibitory interneurons during scaling-driven damage would be beneficial, and could reveal previously-unknown effects on network behaviour and the spread of Alzheimer's pathology. Extension of the model to permit successful learning and retrieval of patterns would enable these additional dynamics to be examined during disease progression, and the comparison of these results with predictions about degradation of recall performance and increased recall time in chapter 4.

For the sake of improving biological plausibility, it would be of great significance to identify a biologically-realistic mechanism for excitotoxicity based on the abstract scaling factor-proportional deletion mechanism employed in this work. This would make it possible to make testable predictions, potentially also leading to the identification of early biomarkers of Alzheimer's disease.

Even without the identification of such a biologically-realistic excitotoxicity mechanism, the predictions made by this thesis regarding information-selectivity during scaling-driven damage progression may still be validated experimentally. For example, an experiment could be designed in which non-fatal signal-blocking toxins are applied to an *in-vitro* neuronal cell culture. The resulting changes in synaptic scaling as the cells compensate for the reduction in activity can be measured directly according to the various methods described by Turrigiano et al. (1998). It has already been shown by Yu et al. (2010) and Crumiller et al. (2011) that their Fourier information measure can be applied to networks of biological neurons, so information contribution measures for each cell in the culture could be obtained using this method and compared with the magnitudes of scaling changes observed in the culture. This would provide a relatively straightforward way to test the validity of the predictions made in this thesis that cells with the lowest information contribution should respond with the strongest compensatory scaling. Further experiments could then be designed to validate the link proposed by Small (2008) between synaptic scaling and excitotoxic cell death in Alzheimer's disease.

The model could be extended to incorporate a reservoir computing-style architecture

with a trainable readout layer. This would enable the model to perform complex tasks such as reinforcement learning for a robotic (real or simulated) arm (Joshi and Maass, 2005; Dura-Bernal et al., 2013), allowing analysis of the effects of Alzheimer's disease pathology in situations which include real-world data and learning of tasks, rather than learning from abstract stimulation.

8.4.1 Possible neurostimulation to prevent Alzheimer's disease

Finally, the predictions made in this thesis raise the intriguing possibility of developing new, more effective, treatments for the underlying pathology of Alzheimer's disease, rather than simply trying to treat the symptoms.

It is proposed that long-term low-intensity neurostimulation, which restores lost activation within the brain due to damage, would reduce the need for compensatory scaling of cells, and would therefore reduce the imbalances between excitatory and inhibitory circuits caused during scaling. This could slow down or even prevent the cascade of damage caused by scaling-driven calcium-mediated excitotoxicity in Alzheimer's disease, thus preventing progression of the disease.

Low-intensity electroprosthetic stimulation has previously been shown in a computational model to restore activity in otherwise-healthy neurons which had suffered a decrease in external activation (Kerr et al., 2011). Recent clinical trials have reported sustained increases in neural activity following continuous (for one year) deep-brain stimulation in patients with probable Alzheimer's disease, with corresponding cognitive improvements (Smith et al., 2012). Other trials using transcranial magnetic stimulation again showed improved cognitive performance after six months of daily treatment (Rabey et al., 2012), and trials using transcranial direct current stimulation also showed positive effects on cognitive performance, with effects persisting for up to four weeks after the end of stimulation (Hansen, 2012). These types of stimulation act by partially depolarising the neural membrane, making it easier for natural synaptic events to trigger an action potential. However, as these trials are still in their early stages, the question of whether these treat-

ments are merely masking the cognitive symptoms, or are actually halting the progression of the disease, remains to be answered.

The predictions made in this thesis therefore make it possible to describe a mechanism by which such low-intensity stimulation may slow or halt the progression of Alzheimer's disease, thereby leading to new and revolutionary treatments for what will surely otherwise become a global first-world epidemic.

LIST OF REFERENCES

L. Aakerlund and R. Hemmingsen. Neural networks as models of psychopathology. *Biol Psychiatry*, 43(7):471–482, 1998.

L.F. Abbott. Lapicque's introduction of the integrate-and-fire model neuron (1907). *Brain Res Bull*, 50(5):303–304, 1999.

M. Abeles, E. Vaadia, and H. Bergman. Firing patterns of single units in the prefrontal cortex and neural network models. *Network: Computation in Neural Systems*, 1(1): 13–25, 1990.

K. Abuhassan, D. Coyle, and L. Maguire. Employing neuronal networks to investigate the pathophysiological basis of abnormal cortical oscillations in alzheimer's disease. In *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2011, page 2065, 2011.

W.S. Anderson and G. Kreiman. Neuroscience: What we cannot model, we do not understand. *Curr Biol*, 21(3):R123–R125, 2011.

W.S. Anderson, P. Kudela, J. Cho, G.K. Bergey, and P.J. Franaszczuk. Studies of stimulus parameters for seizure disruption using neural network simulations. *Biol Cybern*, 97(2): 173–194, 2007.

W.S. Anderson, F. Azhar, P. Kudela, G.K. Bergey, and P.J. Franaszczuk. Epileptic seizures from abnormal networks: Why some seizures defy predictability. *Epilepsy Res*, 2011.

C. Ballatore, M.Y.L. Virginia, and J.Q. Trojanowski. Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nat Rev Neurosci*, 8(9):663–672, 2007.

M. Bartos, I. Vida, M. Frotscher, A. Meyer, H. Monyer, J.R.P. Geiger, and P. Jonas. Fast synaptic inhibition promotes synchronized gamma oscillations in hippocampal interneuron networks. *Proc National Acad Science*, 99(20):13222, 2002.

R.T. Bartus, R.L. Dean 3rd, B. Beer, and A.S. Lippa. The cholinergic hypothesis of geriatric memory dysfunction. *Science*, 217(4558):408, 1982.

M.J. Berridge. Calcium hypothesis of Alzheimer's disease. *Pfluegers Archiv European Journal of Physiology*, 459(3):441–449, 2010.

M.J. Berridge, P. Lipp, and M.D. Bootman. The versatility and universality of calcium signalling. *Nat Rev Mol Cell Biol*, 1(1):11–21, 2000.

B.S. Bhattacharya, D. Coyle, and L.P. Maguire. A thalamo-cortico-thalamic neural mass model to study alpha rhythms in alzheimer's disease. *Neural Networks*, 2011.

E.L. Bienenstock, L.N. Cooper, and P.W. Munro. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *The Journal of Neuroscience*, 2(1):32–48, 1982.

T. Binzegger, R.J. Douglas, and K.A.C. Martin. A quantitative map of the circuit of cat primary visual cortex. *The Journal of Neuroscience*, 24(39):8441–8453, 2004.

D. Boche and J.A.R. Nicoll. Are we getting to grips with Alzheimer's disease at last? *Brain*, 133(5):1297, 2010.

A. Borst and F.E. Theunissen. Information theory and neural coding. *Nat Neurosci*, 2: 947–958, 1999.

Z.I. Botev, J.F. Grotowski, and D.P. Kroese. Kernel density estimation via diffusion. *The Annals of Statistics*, 38(5):2916–2957, 2010.

J.J. Buccafusco and A.V. Terry. Multiple central nervous system targets for eliciting beneficial effects on memory and cognition. *J Pharmacol Exp Ther*, 295(2):438, 2000.

E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci*, 10(3):186–198, 2009.

M.A. Busche, G. Eichhoff, H. Adelsberger, D. Abramowski, K.H. Wiederhold, C. Haass, M. Staufenbiel, A. Konnerth, and O. Garaschuk. Clusters of hyperactive neurons near amyloid plaques in a mouse model of Alzheimer's disease. *Science Signalling*, 321(5896): 1686, 2008.

Y. Cao and S. Grossberg. Stereopsis and 3d surface perception by spiking neurons in laminar cortical circuits: A method for converting neural rate models into spiking models. *Neural Networks*, 2011.

N.T. Carnevale and M.L. Hines. *The NEURON Book*. Cambridge University Press, New York, 2006.

B. Chandler and S. Grossberg. Joining distributed pattern processing and homeostatic plasticity in recurrent on-center off-surround shunting networks: Noise, saturation, short-term memory, synaptic scaling, and BDNF. *Neural Networks*, 2012.

P.S. Churchland and T.J. Sejnowski. *The computational brain*. The MIT press, 1994.

D.E. Clapham. Calcium signaling. *Cell*, 131(6):1047–1058, 2007.

C. Courtney, D. Farrell, R. Gray, R. Hills, L. Lynch, E. Sellwood, S. Edwards, W. Hardyman, J. Raftery, P. Crome, et al. Long-term donepezil treatment in 565 patients with Alzheimer's disease (AD2000): randomised double-blind trial. *Lancet*, 363(9427):2105, 2004.

M. Crumiller, B. Knight, Y. Yu, and E. Kaplan. Estimating the amount of information conveyed by a population of neurons. *Frontiers in Neuroscience*, 5, 2011.

V. Cutsuridis, S. Cobb, and B.P. Graham. Encoding and retrieval in a model of the hippocampal ca1 microcircuit. *Hippocampus*, 20(3):423–446, 2010.

Y. Dan and M. Poo. Spike timing-dependent plasticity of neural circuits. *Neuron*, 44(1): 23–30, 2004.

G.W. Davis. Homeostatic control of neural activity: from phenomenology to molecular design. *Annu Rev Neurosci*, 29:307–323, 2006.

A. Demuro, I. Parker, and G.E. Stutzmann. Calcium signaling and amyloid toxicity in Alzheimer's disease. *The Journal of biological chemistry*, 285:12463–12468, 2010.

M. D'Esposito. From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481):761, 2007.

P.F. Dominey, M. Hoen, and T. Inui. A neurolinguistic model of grammatical construction processing. *J Cogn Neurosci*, 18(12):2088–2107, 2006.

K. Doya, S. Ishii, A. Pouget, and R.P.N. Rao. *Bayesian brain: Probabilistic approaches to neural coding*. MIT Press, 2007.

D.A. Drachman and J. Leavitt. Human memory and the cholinergic system: A relationship to aging? *Arch Neurol*, 30(2):113, 1974.

W. Duch. Computational models of dementia and neurological problems. *Methods Mol Biol*, 401:305–336, November 2007.

S. Dura-Bernal, G.L. Chadderdon, S.A. Neymotin, J.T. Francis, and W.W. Lytton. Towards a real-time interface between a biomimetic model of sensorimotor cortex and a robotic arm. *Pattern Recognition Letters*, 2013.

M.R. Evans. Random dilution in a neural network for biased patterns. *Journal of Physics A: Mathematical and General*, 22:2103, 1989.

S.C. Feinstein and L. Wilson. Inability of tau to properly regulate neuronal microtubule dynamics: a loss-of-function mechanism by which tau might mediate neuronal cell death. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1739(2-3): 268–279, 2005.

P.T. Francis, A.M. Palmer, M. Snape, and G.K. Wilcock. The cholinergic hypothesis of Alzheimer's disease: a review of progress. *Journal of Neurology, Neurosurgery & Psychiatry*, 66(2):137, 1999.

F. Fröhlich, M. Bazhenov, and T.J. Sejnowski. Pathological effect of homeostatic synaptic scaling on network dynamics in diseases of the cortex. *The Journal of Neuroscience*, 28(7):1709–1720, 2008.

W. Gerstner and W. Kistler. *Spiking neuron models: An introduction*. Cambridge University Press New York, NY, USA, 2002.

W. Gerstner and J.L. van Hemmen. Associative memory in a network of 'spiking' neurons. *Network: Computation in Neural Systems*, 3(2):139–164, 1992.

S. Ghosh-Dastidar and H. Adeli. Third Generation Neural Networks: Spiking Neural Networks. *Advances in Computational Intelligence*, pages 167–178, 2009.

D. Golomb, N. Rubin, and H. Sompolinsky. Willshaw model: Associative memory with sparse coding and low firing rates. *Physical Review A*, 41(4):1843–1854, 1990.

B. Graham and D. Willshaw. Improving recall from an associative memory. *Biol Cybern*, 72(4):337–346, 1995.

B. Graham and D. Willshaw. Capacity and information efficiency of the associative net. *Network: Computation in Neural Systems*, 8(1):35–54, 1997.

K. Gurney. *An introduction to neural networks*. UCL Press (Taylor & Francis group), 1997.

N. Hansen. Action mechanisms of transcranial direct current stimulation in alzheimer's disease and memory loss. *Frontiers in Psychiatry*, 3, 2012.

J. Hardy. The amyloid hypothesis for Alzheimer's disease: a critical reappraisal. *J Neurochem*, 110(4):1129–1134, 2009.

J. Hardy and D.J. Selkoe. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science*, 297(5580):353, 2002.

M.E. Hasselmo. Acetylcholine and learning in a cortical associative memory. *Neural Comput*, 5:32–32, 1993.

M.E. Hasselmo. Runaway synaptic modification in models of cortex: Implications for Alzheimer's disease. *Neural Networks*, 7:13–40, 1994.

M.E. Hasselmo. The role of acetylcholine in learning and memory. *Curr Opin Neurobiol*, 16(6):710–715, 2006.

M.E. Hasselmo and H. Eichenbaum. Hippocampal mechanisms for the context-dependent retrieval of episodes. *Neural Networks*, 18(9):1172–1190, 2005.

M.E. Hasselmo, C. Bodelón, and B.P. Wyble. A proposed function for hippocampal theta rhythm: separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Comput*, 14(4):793–817, 2002.

D.O. Hebb. *The organisation of behaviour*. New York: Wiley, 1949.

R. Henson and D.J. Willshaw. Short-term associative memory. In *Proceedings of the INNS World Congress on Neural Networks, 1995*. Citeseer, 1995.

M.L. Hines, T. Morse, M. Migliore, N.T. Carnevale, and G.M. Shepherd. ModelDB: a database to support computational neuroscience. *J Comput Neurosci*, 17(1):7–11, 2004.

R.E. Hoffman and T.H. McGlashan. Neural network models of schizophrenia. *Neuroscientist*, 7(5):441–454, 2001.

J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci U S A*, 79(8):2554, 1982.

J.J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc Natl Acad Sci U S A*, 81(10):3088, 1984.

D. Horn and E. Ruppin. Compensatory mechanisms in an attractor neural network model of schizophrenia. *Neural Comput*, 7(1):182–205, 1995.

D. Horn, E. Ruppin, M. Usher, and M. Herrmann. Neural network modeling of memory deterioration in alzheimer's disease. *Neural Comput*, 5(5):736–749, 1993.

D. Horn, N. Levy, and E. Ruppin. Neuronal-based synaptic compensation: a computational study in Alzheimer's disease. *Neural Comput*, 8(6):1227–1243, 1996.

M.R. Hynd, H.L. Scott, and P.R. Dodd. Glutamate-mediated excitotoxicity and neurodegeneration in Alzheimer's disease. *Neurochem Int*, 45(5):583–595, 2004.

J. Iglesias and A.E.P. Villa. Neuronal cell death and synaptic pruning driven by spike-timing dependent plasticity. In *Artificial Neural Networks–ICANN 2006*, pages 953–962. Springer, 2006.

E.M. Izhikevich. Which model to use for cortical spiking neurons? *Neural Networks, IEEE Transactions on*, 15(5):1063–1070, 2004.

H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, 148, 2001.

G.V.W. Johnson and W.H. Stoothoff. Tau phosphorylation in neuronal cell function and dysfunction. *J Cell Sci*, 117(24):5721, 2004.

P. Joshi and W. Maass. Movement generation with circuits of spiking neurons. *Neural Computation*, 17(8):1715–1738, 2005.

C. Kerr, S. Neymotin, G. Chadderdon, C. Fietkiewicz, J. Francis, and W. Lytton. Electrostimulation as a prosthesis for repair of information flow in a computer model of neocortex. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 19 (6):1–1, 2011.

N. Kopell, C. Boergers, D. Pervouchine, P. Malerba, and A. Tort. Gamma and theta rhythms in biophysical models of hippocampal circuits. *Hippocampal Microcircuits*, pages 423–457, 2010.

M.D. Kopelman. Remote and autobiographical memory, temporal context memory and frontal atrophy in Korsakoff and Alzheimer patients. *Neuropsychologia*, 27(4):437–460, 1989.

F.M. LaFerla, K.N. Green, and S. Oddo. Intracellular amyloid-β in Alzheimer's disease. *Nat Rev Neurosci*, 8(7):499–509, 2007.

K.P. Lamsa, D.M. Kullmann, and M.A. Woodin. Spike-timing dependent plasticity in inhibitory circuits. *Frontiers in Synaptic Neuroscience*, 2, 2010.

S. Lefort, C. Tomm, J.C. Floyd Sarria, and C.C.H. Petersen. The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. *Neuron*, 61 (2):301, 2009.

C.A. Lemere and E. Masliah. Can Alzheimer disease be prevented by amyloid-β immunotherapy? *Nature Reviews Neurology*, 6(2):108–119, 2010.

W.B. Levy. A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus*, 6(6):579–590, 1996.

R. Linsker. Self-organization in a perceptual network. *Computer*, pages 105–117, 1988.

M. Lukosevicius and H. Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, 2009.

W.W. Lytton and M. Stewart. A rule-based firing model for neural networks. *Int. J. Bioelectromagn*, 7:47–50, 2005.

W.W. Lytton and M. Stewart. Rule-based firing for network simulations. *Neurocomputing*, 69(10):1160–1164, 2006.

W.W. Lytton, A. Omurtag, S.A. Neymotin, and M.L. Hines. Just-in-time connectivity for large spiking networks. *Neural Comput*, 20(11):2745–2756, 2008.

W. Maass, T. Natschläger, and H. Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural computation*, 14(11):2531–2560, 2002.

E. Marder and J.M. Goaillard. Variability, compensation and homeostasis in neuron and network function. *Nat Rev Neurosci*, 7(7):563–574, 2006.

E. Marder and A.A. Prinz. Modeling stability in neuron and network function: the role of activity in homeostasis. *Bioessays*, 24(12):1145–1154, 2002.

H. Markram, J. Lübke, M. Frotscher, and B. Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275(5297):213–215, 1997.

J.L. McClelland, B.L. McNaughton, and R.C. O'Reilly. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev*, 102(3):419–457, 1995.

R.J. McEliece, E.C. Posner, E.R. Rodemich, and S. Venkatesh. The capacity of the hopfield associative memory. *Information Theory, IEEE Transactions on*, 33(4):461–482, 1987.

E.D. Menschik and L.H. Finkel. Cholinergic neuromodulation and Alzheimer's disease: from single cells to network simulations. *Disorders of Brain, Behavior, and Cognition: The Neurocomputational Perspective*, page 19, 1999.

L. Minati, T. Edginton, M. Grazia Bruzzone, and G. Giaccone. Reviews: Current Concepts in Alzheimer's Disease: A Multidisciplinary Review. *American Journal of Alzheimers Disease and Other Dementias*, 24(2):95, 2009.

A. Morse and M. Aktius. Dynamic liquid association: Complex learning without implausible guidance. *Neural Networks*, 22(7):875–889, 2009.

T.M. Morse, N.T. Carnevale, P.G. Mutalik, M. Migliore, and G.M. Shepherd. Abnormal excitability of oblique dendrites implicated in early alzheimer's: A computational study. *Frontiers in Neural Circuits*, 4, 2010.

C. Mount and C. Downton. Alzheimer disease: progress or profit? *Nat Med*, 12(7): 780–784, 2006.

J. Movellan. Contrastive Hebbian learning in the continuous Hopfield model. In *Proceedings of the 1990 Connectionist Models Summer School*, pages 10–17, 1990.

S.A. Neymotin, K.M. Jacobs, A.A. Fenton, and W.W. Lytton. Synaptic information transfer in computer models of neocortical columns. *J Comput Neurosci*, 30(1):69–84, 2011a.

S.A. Neymotin, C.C. Kerr, J.T. Francis, and W.W. Lytton. Training oscillatory dynamics with spike-timing-dependent plasticity in a computer model of neocortex. In *Signal Processing in Medicine and Biology Symposium (SPMB)*, pages 1–6. IEEE, 2011b.

S.A. Neymotin, H. Lee, E. Park, A.A. Fenton, and W.W. Lytton. Emergence of physiological oscillation frequencies in a computer model of neocortex. *Frontiers in Computational Neuroscience*, 5, 2011c.

D.W. Nicholson. Good and bad cell death. *Nature*, 457(7232):970–971, 2009.

A. Nikolaev, T. McLaughlin, D. O'Leary, and M. Tessier-Lavigne. N-APP binds DR6 to cause axon pruning and neuron death via distinct caspases. *Nature*, 457(7232):981, 2009.

E. Oja. Simplified neuron model as a principal component analyzer. *Journal of mathematical biology*, 15(3):267–273, 1982.

R.C. O'Reilly. *The LEABRA model of neural interactions and learning in the neocortex*. PhD thesis, Carnegie Mellon University, 1996.

R.C. O'Reilly. Generalization in interactive networks: The benefits of inhibitory competition and Hebbian learning. *Neural Comput*, 13(6):1199–1241, 2001.

R.C. O'Reilly and Y. Munakata. *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. The MIT Press, 2000.

R.C. O'Reilly and J.W. Rudy. Computational principles of learning in the neocortex and hippocampus. *Hippocampus*, 10(4):389–397, 2000.

R.C. O'Reilly, T.S. Braver, and J.D. Cohen. A biologically based computational model of working memory. *Models of working memory: Mechanisms of active maintenance and executive control*, pages 375–411, 1999.

L. Paninski. Estimation of entropy and mutual information. *Neural Comput*, 15(6): 1191–1253, 2003.

S. Panzeri and A. Treves. Analytical estimates of limited sampling biases in different information measures. *Network-Computation in Neural Systems*, 7(1):87–108, 1996.

G.A. Prieto, R.L. Parker, and F.L. Vernon III. A Fortran 90 library for multitaper spectrum analysis. *Computers & Geosciences*, 35(8):1701–1710, 2009.

R.Q. Quiroga and S. Panzeri. Extracting information from neuronal populations: information theory and decoding approaches. *Nat Rev Neurosci*, 10(3):173–185, 2009.

J.M. Rabey, E. Dobronevsky, S. Aichenbaum, O. Gonen, R.G. Marton, and M. Khaigrekht. Repetitive transcranial magnetic stimulation combined with cognitive training is a safe and effective modality for the treatment of alzheimer's disease: a randomized, double-blind study. *Journal of Neural Transmission*, pages 1–7, 2012.

R. Reed. Pruning algorithms – a survey. *IEEE Transactions on Neural Networks*, 4(5): 740–747, September 1993.

R.F. Reinhart and J.J. Steil. Attractor-based computation with reservoirs for online learning of inverse kinematics. *ESANN, April*, 2009:257–262, 2009.

B.H. Ridha, J. Barnes, J.W. Bartlett, A. Godbolt, T. Pepple, M.N. Rossor, and N.C. Fox. Tracking atrophy progression in familial Alzheimer's disease: a serial MRI study. *The Lancet Neurology*, 5(10):828–834, 2006.

E.T. Rolls. A computational theory of episodic memory formation in the hippocampus. *Behav Brain Res*, 215(2):180–196, 2010.

E.T. Rolls and R.P. Kesner. A computational theory of hippocampal function, and empirical tests of the theory. *Prog Neurobiol*, 79(1):1–48, 2006.

E.T. Rolls, A. Treves, and M.J. Tovee. The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. *Exp Brain Res*, 114(1):149–162, 1997.

M. Rowan. Effects of Compensation, Connectivity and Tau in a Computational Model of Alzheimer's Disease. In *The 2011 International Joint Conference on Neural Networks (IJCNN)*, pages 543–550. IEEE, 2011.

M. Rowan. Information-selectivity of beta-amyloid pathology in an associative memory model. *Frontiers in Computational Neuroscience*, 6(2), January 2012.

M. Rowan and S. Neymotin. Synaptic scaling balances learning in a spiking model of neocortex. In M. Tomassini et al., editor, *11th International Conference on Adaptive and Natural Computing Algorithms (ICANNGA)*, volume 7824 of *Lecture Notes in Computer Science*, pages 20–29, Lausanne, Switzerland, 2013. Springer-Verlag.

E. Ruppin. Neural modelling of psychiatric disorders. *Network: Computation in Neural Systems*, 6(4):635–656, 1995.

E. Ruppin and J.A. Reggia. A neural model of memory impairment in diffuse cerebral atrophy. *The British Journal of Psychiatry*, 166(1):19–28, 1995a.

E. Ruppin and J.A. Reggia. Patterns of functional damage in neural network models of associative memory. *Neural Comput*, 7(5):1105–1127, 1995b.

L.C. Rutherford, S.B. Nelson, and G.G. Turrigiano. BDNF has opposite effects on the quantal amplitude of pyramidal neuron and interneuron excitatory synapses. *Neuron*, 21(3):521–530, 1998.

T.D. Sanger. Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural networks*, 2(6):459–473, 1989.

A. Savioz, G. Leuba, P.G. Vallet, and C. Walzer. Contribution of neural networks to Alzheimer disease's progression. *Brain Res Bull*, 2009.

U.D. Schiller and J.J. Steil. Analyzing the weight dynamics of recurrent learning algorithms. *Neurocomputing*, 63:5–23, 2005.

B. Schrauwen, D. Verstraeten, and J. van Campenhout. An overview of reservoir computing: theory, applications and implementations. In *Proceedings of the 15th European Symposium on Artificial Neural Networks*, pages 471–482, 2007.

W.E. Skaggs, B.L. McNaughton, K.M. Gothard, and E.J. Markus. An information-theoretic approach to deciphering the hippocampal code. In *Advances in Neural Information Processing Systems 5, [NIPS Conference]*, pages 1030–1037, San Francisco, CA, USA, 1993. Morgan Kaufmann Publishers Inc.

D.H. Small. Network dysfunction in Alzheimer's disease: does synaptic scaling drive disease progression? *Trends Mol Med*, 14(3):103 – 108, 2008.

D.H. Small. Dysregulation of calcium homeostasis in alzheimer's disease. *Neurochem Res*, 34(10):1824–1829, 2009.

G.S. Smith, A.W. Laxton, D.F. Tang-Wai, M.P. McAndrews, A.O. Diaconescu, C.I. Workman, and A.M. Lozano. Increased cerebral metabolism after 1 year of deep brain stimulation in alzheimer disease. *Archives of Neurology*, pages archneurol–2012, 2012.

T.L. Spires-Jones, W.H. Stoothoff, A. de Calignon, P.B. Jones, and B.T. Hyman. Tau pathophysiology in neurodegeneration: a tangled issue. *Trends Neurosci*, 32(3):150–159, 2009.

M. Spitzer. The history of neural network research in psychopathology. *Neural Networks and Psychopathology: Connectionist Models in Practice and Research*, page 14, 1998.

W.C. Stacey, M.T. Lazarewicz, and B. Litt. Synaptic noise and physiological coupling generate high-frequency oscillations in a hippocampal computational model. *J Neurophysiol*, 102(4):2342–2357, 2009.

J.J. Steil. Backpropagation-Decorrelation: Online recurrent learning with O (N) complexity. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, volume 1, pages 843–848, 2004.

J.J. Steil. Online reservoir adaptation by intrinsic plasticity for backpropagation-decorrelation and echo state learning. *Neural Networks*, 20(3):353–364, 2007.

D.J. Stein and J. Ludik. *Neural networks and psychopathology: connectionist models in practice and research*. Cambridge University Press, 1998.

K. Supekar, V. Menon, D. Rubin, M. Musen, and M.D. Greicius. Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. *PLoS Comput Biol*, 4(6): e1000100, 2008.

A.V. Terry and J.J. Buccafusco. The cholinergic hypothesis of age and Alzheimer's disease-related cognitive deficits: recent challenges and their implications for novel drug development. *J Pharmacol Exp Ther*, 306(3):821, 2003.

P. Tiraboschi, L. A. Hansen, L. J. Thal, and J. Corey-Bloom. The importance of neuritic plaques and tangles to the development and evolution of AD. *Neurology*, 62(11):1984–9, June 2004.

A.B.L. Tort, H.G. Rotstein, T. Dugladze, T. Gloveli, and N.J. Kopell. On the formation of gamma-coherent cell assemblies by oriens lacunosum-moleculare interneurons in the hippocampus. *Proc National Acad Science*, 104(33):13490, 2007.

C.A. Trasande and J.M. Ramirez. Activity deprivation leads to seizures in hippocampal slice cultures: is epilepsy the consequence of homeostatic plasticity? *J Clin Neurophysiol*, 24(2):154–164, 2007.

R.D. Traub, D. Contreras, M.O. Cunningham, H. Murray, F.E.N. LeBeau, A. Roopun, A. Bibbig, W.B. Wilent, M.J. Higley, and M.A. Whittington. Single-column thalamocortical network model exhibiting gamma oscillations, sleep spindles, and epileptogenic bursts. *J Neurophysiol*, 93(4):2194, 2005.

A. Treves and E.T. Rolls. What determines the capacity of autoassociative memories in the brain? *Network: Computation in Neural Systems*, 2(4):371–397, 1991.

M.V. Tsodyks and M.V. Feigel'Man. The enhanced storage capacity in neural networks with low activity level. *EPL (Europhysics Letters)*, 6:101–105, 1988.

G. Turrigiano. Too many cooks? intrinsic and synaptic homeostatic mechanisms in cortical circuit refinement. *Annu Rev Neurosci*, 34:89–103, 2011.

G.G. Turrigiano. The self-tuning neuron: synaptic scaling of excitatory synapses. *Cell*, 135(3):422–435, 2008.

G.G. Turrigiano, K.R. Leslie, N.S. Desai, L.C. Rutherford, S.B. Nelson, et al. Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, pages 892–895, 1998.

M.C.W. van Rossum, G.Q. Bi, and G.G. Turrigiano. Stable Hebbian learning from spike timing-dependent plasticity. *J Neurosci*, 20(23):8812, 2000.

D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt. An experimental unification of reservoir computing methods. *Neural Networks*, 20(3):391–403, 2007.

D. Vierling-Claassen, J.A. Cardin, C.I. Moore, and S.R. Jones. Computational modeling of distinct neocortical oscillations driven by cell-type selective optogenetic drive: separable resonant circuits controlled by low-threshold spiking and fast-spiking interneurons. *Frontiers in Human Neuroscience*, 4, 2010.

G.V. Wallenstein and M.E. Hasselmo. Are there common neural mechanisms for learning, epilepsy, and Alzheimer's disease? *Neural Networks and Psychopathology: Connectionist Models in Practice and Research*, pages 316–344, 1998.

D.J. Watts and S.H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393 (6684):440–442, 1998.

D.M. Wilcock, N. Gharkholonarehe, W.E. van Nostrand, J. Davis, M.P. Vitek, and C.A. Colton. Amyloid Reduction by Amyloid-β Vaccination Also Reduces Mouse Tau Pathology and Protects from Neuron Loss in Two Mouse Models of Alzheimer's Disease. *J Neurosci*, 29(25):7957, 2009.

D.J. Willshaw, O.P. Buneman, and H.C. Longuet-Higgins. Non-holographic associative memory. *Nature*, 222(5197):960–962, 1969.

T. Yamazaki and S. Tanaka. The cerebellum as a liquid state machine. *Neural Networks*, 20(3):290–297, 2007.

Y. Yu, M. Crumiller, B. Knight, and E. Kaplan. Estimating the amount of information carried by a neuronal population. *Frontiers in computational neuroscience*, 4, 2010.

L. Zhang, H. Tao, C. Holt, W. Harris, and M. Poo. A critical window for cooperation and competition among developing retinotectal synapses. *Nature*, 395(6697):37–44, 1998.